

SYNTAKTISCHE UND REFERENZIELLE  
SPRACHVERSTEHENSPROZESSE  
AUS KONNEKTIONISTISCHER PERSPEKTIVE

Inaugural-Dissertation  
zur  
Erlangung der Doktorwürde  
der Wirtschafts- und Verhaltenswissenschaftlichen Fakultät  
der Albert-Ludwigs-Universität Freiburg i. Br.

vorgelegt von

**Helmut Weldle**  
aus Oberndorf a. N.

SS 2011

---

Dekan: Prof. Dr. Dieter K. Tscheulin

Referent: PD Dr. Lars Konieczny

Korreferent: PD Dr. Christoph Hölscher

Korreferentin: Prof. Dr. Evelyn Ferstl

Datum des Promotionsbeschlusses: 05. Oktober 2011

# Inhaltsverzeichnis

<b>Tabellenverzeichnis</b>	<b>vii</b>
<b>Abbildungsverzeichnis</b>	<b>ix</b>
<b>Vorwort</b>	<b>xiii</b>
<b>1. Einleitung</b>	<b>1</b>
1.1. Sprache, Referenz, Kognition . . . . .	1
1.2. Ausgangspunkte und Projekte der Arbeit . . . . .	4
1.3. Übersicht zur Gliederung der Arbeit . . . . .	7
<b>I. Konnektionismus und Sprache</b>	<b>11</b>
<b>2. Konnektionismus</b>	<b>13</b>
2.1. Kognitive Modellierung und Konnektionismus . . . . .	13
2.1.1. Gütekriterien und Ziele der Modellierung . . . . .	14
2.1.2. Einordnung konnektionistischer Modellierung . . . . .	15
2.2. Konnektionismus und Kognition . . . . .	17
2.2.1. Parallele Verarbeitung . . . . .	18
2.2.2. Verteilte Repräsentation . . . . .	24
2.3. Konnektionismus und Sprache . . . . .	28
<b>3. Sprachverarbeitung in Simple Recurrent Networks</b>	<b>33</b>
3.1. Sequenzielle Verarbeitung: Modellierung mit SRNs . . . . .	33
3.1.1. Alternative rekurrente Architekturen . . . . .	35
3.1.2. Sequenzerfassung durch Prädiktion . . . . .	36
3.1.3. Eigenschaften der Sprachverarbeitung in SRNs . . . . .	39
3.1.4. Die Frage der Systematizität der Verarbeitung . . . . .	51
3.2. SRNs als psycholinguistisches Modell . . . . .	59
3.2.1. Beschränkte Kapazität, Interferenz und Antizipation . . . . .	60
3.2.2. Frequenz, Regularität, Erfahrung . . . . .	70
3.2.3. Sensitivität für den lokalen Kontext . . . . .	75
3.2.4. SRNs als probabilistische Parser . . . . .	78
<b>4. Konnektionistische Modelle des Sprachverstehens</b>	<b>83</b>
4.1. Modelle und Entwicklungslinien . . . . .	84
4.2. Bedeutungskonstruktion durch Abfragemechanismen . . . . .	86
4.2.1. Sentence/Message Gestalt . . . . .	86

4.2.2.	Connectionist Sentence Comprehension and Production . . . . .	89
4.3.	SOM-gestützte Strukturabbildung: INSOMNet . . . . .	93
4.4.	Direkte Assoziation und Integration . . . . .	96
4.4.1.	Direkte Assoziation: Distributed Situation Space . . . . .	97
4.4.2.	Assoziation und Integration: Coordinated Interplay Account . . . . .	101
4.5.	Konzeption semantischer Verarbeitung . . . . .	104

## **II. Koreferenz und lokale Kohärenz 109**

### **5. Anaphernauffösung und lokale syntaktische Kohärenzen 111**

5.1.	Satzverarbeitung: Perspektiven und Modelle . . . . .	111
5.2.	Lokale syntaktische Kohärenzen . . . . .	117
5.2.1.	Einfluss auf die Satzverarbeitung: Interferenz . . . . .	118
5.2.2.	Einfluss auf die Satzverarbeitung: Interpretation . . . . .	120
5.2.3.	Probleme und offene Fragen . . . . .	127
5.2.4.	Theoretische Erklärungsansätze . . . . .	131
5.3.	Anaphernauffösung . . . . .	138
5.3.1.	Strukturelle Ansätze: Bindungsdomänen . . . . .	139
5.3.2.	Online-Anaphernauffösung: Empirische Befunde . . . . .	143

### **6. AnaSRN: SRNs und Referenzialität 151**

6.1.	Konnektionistische Modellierung der Anaphernauffösung . . . . .	152
6.1.1.	Verarbeitung struktureller Bindungs-Constraints durch SRNs . . . . .	153
6.1.2.	Kritik an R. Frank et al. (2005) . . . . .	158
6.2.	Anaphernauffösung in SRNs . . . . .	160
6.2.1.	Integratives holistisches Training . . . . .	160
6.2.2.	Architektur mit kaskadierenden SRNs . . . . .	162
6.2.3.	AnaSRN: Simulationsstudien . . . . .	164
6.3.	Simulationsstudie 1 . . . . .	171
6.3.1.	Trainings- und Test-Stimuli . . . . .	171
6.3.2.	Technische Details zur Simulation . . . . .	173
6.3.3.	Ergebnisse . . . . .	174
6.3.4.	Diskussion . . . . .	194
6.4.	Simulationsstudie 2 . . . . .	197
6.4.1.	Trainings- und Test-Stimuli . . . . .	198
6.4.2.	Technische Details zur Simulation . . . . .	199
6.4.3.	Ergebnisse . . . . .	200
6.4.4.	Diskussion . . . . .	217
6.5.	Simulationsstudie 3: Studien mit reduzierten Stimuli . . . . .	219
6.5.1.	Korpus 3a: Effekte fehlender Satzstruktur-Markierung . . . . .	220
6.5.2.	Korpus 3b: Vorenthaltener Referent . . . . .	232
6.5.3.	Korpus 3c: Vergleichsstudie ohne Referenzzuweisung . . . . .	238
6.6.	Diskussion des Modells AnaSRN . . . . .	240
6.6.1.	Ableitung empirischer Hypothesen . . . . .	240
6.6.2.	Kritische Evaluation des Modells . . . . .	242

---

<b>7. AnaLoCo: Anaphern und lokale Kohärenzen</b>	<b>245</b>
7.1. Temporäre lokale Bindungsdomänen . . . . .	245
7.1.1. Gegenstand und Fragestellung . . . . .	246
7.1.2. Empirische Untersuchung zur Anaphernverarbeitung in LSK . . .	247
7.2. AnaLoCo 1: Lesezeitmessung . . . . .	249
7.2.1. Design . . . . .	250
7.2.2. Stimulus-Material . . . . .	250
7.2.3. Methodik und Hypothesen . . . . .	252
7.2.4. Ergebnisse . . . . .	255
7.2.5. Diskussion AnaLoCo 1 . . . . .	267
7.3. AnaLoCo 2: Visual-World-Experiment mit Referenten . . . . .	271
7.3.1. Design . . . . .	272
7.3.2. Stimulus-Materialien . . . . .	272
7.3.3. Methodik und Hypothesen . . . . .	276
7.3.4. Ergebnisse . . . . .	279
7.3.5. Diskussion AnaLoCo 2 . . . . .	288
7.3.6. Auswirkung prosodischer Manipulationen . . . . .	290
7.4. AnaLoCo 3: Visual-World-Experiment mit Szenen . . . . .	296
7.4.1. Design . . . . .	297
7.4.2. Stimulus-Materialien . . . . .	298
7.4.3. Methodik und Hypothesen . . . . .	302
7.4.4. Ergebnisse . . . . .	305
7.4.5. Diskussion AnaLoCo 3 . . . . .	312
7.5. Diskussion der Befunde und Konsequenzen . . . . .	316
7.5.1. Zusammenfassung zentraler Ergebnisse . . . . .	317
7.5.2. Bezug zu Theorien für LSK und Anaphernauflösung . . . . .	319
<b>III. Konnektionismus, Embodiment und Sprachverstehen</b>	<b>325</b>
<b>8. Embodied Cognition und Sprachverarbeitung</b>	<b>327</b>
8.1. Embodied Cognition . . . . .	328
8.1.1. Theoretische Grundlagen . . . . .	329
8.1.2. Konzeptualisierung und Kognition . . . . .	332
8.2. Sprachverarbeitung im Embodiment-Ansatz . . . . .	337
8.2.1. Linguistische Indizierung von Konzepten . . . . .	338
8.2.2. Linguistisch kontrollierte Simulation . . . . .	339
8.2.3. Perspektivische Situationssimulation . . . . .	341
8.2.4. Embodied Construction Grammar . . . . .	343
8.3. Empirische Befunde zum Embodiment-Ansatz . . . . .	344
8.3.1. Action-Sentence Compatibility Effect . . . . .	345
8.3.2. Effekte der Perzeption: Form, Farbe, Orientierung . . . . .	347
8.3.3. Neuropsychologische Befunde . . . . .	349
8.3.4. Perspektivische Simulationen in der Satzverarbeitung . . . . .	351

<b>9. Embodied Cognition und Konnektionismus</b>	<b>357</b>
9.1. Modellierungen von Embodied Language Processing . . . . .	358
9.1.1. Integration linguistischer und visueller Verarbeitung . . . . .	358
9.1.2. Erwerb und Verarbeitung sensomotorisch verankerter Sprache . . . . .	361
9.1.3. Weitere Ansätze zum Sprach- und Bedeutungserwerb . . . . .	362
9.1.4. Sprachgenese in Robotern und autonomen Agenten . . . . .	365
9.2. Eine gemeinsame Perspektive . . . . .	367
9.2.1. Gemeinsamkeiten und Differenzen . . . . .	367
9.2.2. Chancen und Nutzen einer gemeinsamen Perspektive . . . . .	370
9.3. Konnektionismus und Embodied Language Processing . . . . .	374
9.3.1. Modelle lexikalischer Verarbeitung . . . . .	375
9.3.2. Modelle syntaktischer Verarbeitung . . . . .	383
9.3.3. Kriterien für Modelle des Satzverstehens . . . . .	390
<b>10. Konnektionistische Modellierung: Sprachemergenz und Embodiment</b>	<b>393</b>
10.1. Modellierungsansatz des Projekts SynSem . . . . .	394
10.1.1. Architektur und Aufgaben . . . . .	394
10.1.2. Gestaltung der Stimuli . . . . .	398
10.1.3. Aufgaben-Spektrum und Integration . . . . .	402
10.2. Modellierung SynSem I . . . . .	406
10.2.1. Stimuli und Simulationen . . . . .	406
10.2.2. Auswertungsverfahren . . . . .	408
10.2.3. Ergebnisse . . . . .	411
10.2.4. Diskussion von SynSem I . . . . .	430
10.3. Modellierung SynSem II . . . . .	434
10.3.1. Stimuli und neue Architektur . . . . .	435
10.3.2. Simulationen . . . . .	440
10.3.3. Ergebnisse Variante A: Positionsrauschen . . . . .	442
10.3.4. Ergebnisse Variante B: Relationsrauschen . . . . .	453
10.3.5. Ergebnisse Variante C: Systematizität . . . . .	459
10.3.6. Ergebnisse zur Sprachproduktion . . . . .	461
10.3.7. Diskussion von SynSem 2 . . . . .	464
10.4. Perspektive von SynSem auf Sprachverarbeitung . . . . .	468
10.4.1. Modellierungskriterien . . . . .	468
10.4.2. Eigenschaften des Modells . . . . .	469
10.4.3. Bedeutung im SynSem-Ansatz . . . . .	471
10.4.4. Weiterentwicklungsmöglichkeiten für SynSem . . . . .	473
<b>11. Schluss</b>	<b>477</b>
11.1. Untersuchungsbereiche und Argumentationslinien . . . . .	477
11.2. Zentrale Ergebnisse und Schlussfolgerungen . . . . .	478
11.3. Perspektiven . . . . .	482
<b>Literaturverzeichnis</b>	<b>485</b>

# Tabellenverzeichnis

5.1. Theoretische Ansätze für Bindungs-Constraints und -Präferenzen . . . . .	139
6.1. Grundzüge der PCFG 1 . . . . .	166
6.2. Grundzüge der PCFG 2 . . . . .	166
6.3. Erläuterung der Konstituenten zur PCFG . . . . .	167
6.4. Lexikon 1 zur PCFG . . . . .	168
6.5. Lexikon 2 zur PCFG . . . . .	168
7.1. Design Experiment 1 . . . . .	250
7.2. LMER AnaLoCo Exp. 1, Pos.: Präpositionalphrase . . . . .	264
7.3. LMER AnaLoCo Exp. 1, Pos.: Finites Matrixsatz-Verb . . . . .	265
7.4. LMER AnaLoCo Exp. 1, Pos.: Beginn des nachfolgenden Hauptsatzes . . . . .	266
7.5. Logit Mixed Model AnaLoCo Exp. 1, Korrektheit der Antworten . . . . .	267
7.6. LMER AnaLoCo Exp. 1, Antwortlatenzen . . . . .	268
7.7. Design Experiment 2 . . . . .	272
7.8. LMER AnaLoCo Exp. 2, Bed. Reflexivpronomen (1000 ms) . . . . .	286
7.9. LMER AnaLoCo Exp. 2, Bed. Personalpronomen (600 ms), lokale IA . . . . .	287
7.10. LMER AnaLoCo Exp. 2, Bed. Personalpronomen (600 ms), lokale IA . . . . .	287
7.11. Design Experiment 3 . . . . .	297
7.12. LMER AnaLoCo Exp. 3, Bed. Reflexivpronomen (1000ms) . . . . .	310
7.13. LMER AnaLoCo Exp. 3, Bed. Personalpronomen (800ms), lokale IA . . . . .	311
7.14. LMER AnaLoCo Exp. 3, Bed. Reflexivpronomen (1000ms), lokale IA . . . . .	312
7.15. LMER AnaLoCo Exp. 3, Bed. Personalpronomen (800ms), lokale IA . . . . .	313



# Abbildungsverzeichnis

2.1. Architektur: FFN . . . . .	20
3.1. Architektur: SRN . . . . .	34
4.1. Architektur: SG-Modell . . . . .	88
4.2. Architektur: CSCP . . . . .	90
4.3. Architektur: INSOMNet . . . . .	94
4.4. Architektur: DSS-Modell . . . . .	98
4.5. Architektur: CIA-Net . . . . .	102
5.1. Strukturelle Constraints für Reflexiv- und Personalpronomen . . . . .	141
5.2. Strukturelle Constraints für Reflexivpronomen in komplexen Sätzen . . . . .	142
5.3. Beschränkung struktureller Constraints auf die lokale Satzebene . . . . .	142
6.1. Architektur: Kaskadierende SRNs . . . . .	163
6.2. AnaSRN 1: GPEn für Nebensatz mit Subjekt-Relativsatz . . . . .	177
6.3. AnaSRN 1: GPEn für Nebensatz mit Objekt-Relativsatz . . . . .	177
6.4. AnaSRN 1: Aktivierung Position Matrix-Verb NSSRS . . . . .	178
6.5. AnaSRN 1: Aktivierung Position Reflexiv-/Personalpronomen NSSRS . . . . .	179
6.6. AnaSRN 1: GPEn für Hauptsatz mit Subjekt-Relativsatz . . . . .	180
6.7. AnaSRN 1: GPEn für Hauptsatz mit Objekt-Relativsatz . . . . .	181
6.8. AnaSRN 1: Entwicklung des RAEn für Reflexivpronomen . . . . .	183
6.9. AnaSRN 1: Entwicklung des RAEn für Personalpronomen . . . . .	183
6.10. AnaSRN 1: Aktivierung für Reflexivpronomen . . . . .	186
6.11. AnaSRN 1: Aktivierung für Personalpronomen . . . . .	189
6.12. AnaSRN 1: HCA für Hidden-Layer Wortvorhersage nach Wort . . . . .	191
6.13. AnaSRN 1: HCA für Hidden-Layer Referenzzuweisung nach Wort . . . . .	192
6.14. AnaSRN 1: HCA für Hidden-Layer Wortvorhersage nach Satztyp . . . . .	193
6.15. AnaSRN 1: HCA für Hidden-Layer Referenzzuweisung nach Satztyp . . . . .	194
6.16. AnaSRN 2: GPEn für NSORS . . . . .	201
6.17. AnaSRN 2: Aktivierung für anaphorische Referenz in HSORS . . . . .	201
6.18. AnaSRN 2: Aktivierung für das Matrixsatz-Verb in NSORS . . . . .	202
6.19. AnaSRN 2: GPEn für HSSRS . . . . .	203
6.20. AnaSRN 2: Aktivierung für das Matrixsatz-Verb in HSSRS . . . . .	204
6.21. AnaSRN 2: Aktivierung für anaphorische Referenz in HSORS . . . . .	205
6.22. AnaSRN 2: Vgl. RAEn, Reflexivpronomen über verschiedene Satztypen . . . . .	206
6.23. AnaSRN 2: Vgl. RAEn, Personalpronomen über verschiedene Satztypen . . . . .	207
6.24. AnaSRN 2: Vgl. RAEn für verschiedene anaphorische Formen . . . . .	208
6.25. AnaSRN 2: Aktivierung für Reflexivpronomen nach Referentenklassen . . . . .	211

---

6.26.	AnaSRN 2: Aktivierung für Reflexivpronomen nach Referentenklassen	212
6.27.	AnaSRN 2: Aktivierung für Personalpronomen nach Referentenklassen	214
6.28.	AnaSRN 2: HCA für Hidden-Layer Wortvorhersage nach Wort	216
6.29.	AnaSRN 2: HCA für Hidden-Layer Wortvorhersage nach Satztyp	216
6.30.	AnaSRN 2: HCA für Hidden-Layer Referenzzuweisung nach Satztyp	217
6.31.	AnaSRN 2: HCA für Hidden-Layer Referenzzuweisung nach Wort	218
6.32.	AnaSRN 3a: GPEN für NSORS	222
6.33.	AnaSRN 3a: GPEN für NSORS(Kontrollbedingung)	222
6.34.	AnaSRN 3a: GPEN für HSORS	223
6.35.	AnaSRN 3a: Aktivierung für Anapher in NSORS	224
6.36.	AnaSRN 3a: Aktivierung für Matrixsatz-Verb in HSORS	224
6.37.	AnaSRN 3a: Aktivierung für Matrixsatz-Verb in NSORS	225
6.38.	AnaSRN 3a: Aktivierung für Anapher in HSORS	226
6.39.	AnaSRN 3a: Vgl. RAEn, Reflexivpronomen über verschiedene Satztypen	227
6.40.	AnaSRN 3a: Vgl. RAEn, Personalpronomen über verschiedene Satztypen	228
6.41.	AnaSRN 3a: Vgl. RAEn, Reflexivpronomen über verschiedene Satztypen	228
6.42.	AnaSRN 3a: Aktivierung für Reflexivpronomen nach Referentenklassen	230
6.43.	AnaSRN 3a: Aktivierung für Personalpronomen nach Referentenklassen	231
6.44.	AnaSRN 3b: RAEn für Reflexivpronomen in einfachem HS	234
6.45.	AnaSRN 3c: HCA für Hidden-Layer nach Satztyp	239
7.1.	Ergebnisse Exp. 1: Alle Bedingungen	258
7.2.	Ergebnisse Exp. 1: LSK	258
7.3.	Ergebnisse Exp. 1: Anapher	259
7.4.	Ergebnisse Exp. 1: Anteil korrekt beantworteter Fragen	260
7.5.	Ergebnisse Exp. 1: Latenzen der Beantwortung	261
7.6.	Visuelle Stimuli Experiment 2	275
7.7.	Ergebnisse Exp. 2: Bedingung Reflexivpronomen	282
7.8.	Ergebnisse Exp. 2: Bedingung Personalpronomen	283
7.9.	Ergebnisse Exp. 2: Balkendiagramm für Bedingung Reflexivpronomen	285
7.10.	Ergebnisse Exp. 2: Balkendiagramm für Bedingung Personalpronomen	285
7.11.	Ergebnisse Rating-Studie: Verständlichkeit	293
7.12.	Ergebnisse Rating-Studie: Prosodie	294
7.13.	Visuelle Stimuli Experiment 3: Reflexive Szenen	301
7.14.	Visuelle Stimuli Experiment 3: Transitive Szenen	301
7.15.	Ergebnisse Exp. 3: Bedingung Reflexivpronomen	306
7.16.	Ergebnisse Exp. 3: Bedingung Personalpronomen	307
7.17.	Ergebnisse Exp. 3: Balkendiagramm für Bedingung Reflexivpronomen	309
7.18.	Ergebnisse Exp. 3: Balkendiagramm für Bedingung Personalpronomen	309
8.1.	Immersed Experiencer Framework	342
9.1.	Architektur: Dual-Route-Netzwerk	376
10.1.	Architektur: Dual-Route-SRN	395
10.2.	Stimuli SynSem I	398
10.3.	SynSem I: Beispiel (entkoppeltes) Verstehen	412

---

10.4.	SynSem I: Muster-Assoziation bei entkoppeltem Verstehen . . . . .	414
10.5.	SynSem I: Identifizierung/Lokalisierung, entkoppeltes Verstehen . . .	415
10.6.	SynSem I: Wortvorhersage bei entkoppeltem Verstehen . . . . .	416
10.7.	SynSem I: Beispiel 1 zum situierten Verstehen . . . . .	418
10.8.	SynSem I: Beispiel 2 zum situierten Verstehen . . . . .	420
10.9.	SynSem I: Muster-Assoziation bei situierter Sprachverarbeitung . . . .	422
10.10.	SynSem I: Objekt-Identifizierung bei situierter Sprachverarbeitung . .	423
10.11.	SynSem I: Objekt-Lokalisierung bei situierter Sprachverarbeitung . . .	424
10.12.	SynSem I: Wortvorhersage bei situierter Sprachverarbeitung . . . . .	425
10.13.	SynSem I: Beispiel zur beschreibenden Produktion . . . . .	427
10.14.	SynSem I: Wortvorhersage Produktion . . . . .	429
10.15.	SynSem I: Performanz der visuellen Route bei Satz-Produktion . . . .	431
10.16.	Stimuli SynSem II . . . . .	436
10.17.	Architektur: Retinale Bereiche SynSem II . . . . .	437
10.18.	Architektur: Rezeptive Felder SynSem II . . . . .	439
10.19.	SynSem II, Variante A: Visuellen Route, entkoppeltes Verstehen . . . .	445
10.20.	SynSem II, Variante A: Wortvorhersage, entkoppeltes Verstehen . . . .	446
10.21.	SynSem II, Variante A.a: Beispiel situiertes Verstehen . . . . .	448
10.22.	SynSem II, Variante A.b: Beispiel situiertes Verstehen . . . . .	449
10.23.	SynSem II, Variante A: Muster-Assoziation, situierte Sprachverarbeitung	451
10.24.	SynSem II, Variante A: Wortvorhersage, situierte Verarbeitung . . . . .	452
10.25.	SynSem II, Variante A: Wortvorhersage für Fragesätze . . . . .	452
10.26.	SynSem II, Variante B.c: Beispiel entkoppeltes Verstehen . . . . .	455
10.27.	SynSem II, Variante B: Visuellen Route, entkoppeltes Verstehen . . . .	458
10.28.	SynSem II, Variante B: Wortvorhersage, entkoppeltes Verstehen . . . .	459
10.29.	SynSem II, Variante B.c: Beispiel Sprachproduktion . . . . .	463
10.30.	SynSem II: Beschreibende Sprachproduktion, Varianten A, B und C . .	465



# Vorwort

Die vorliegende Dissertation beschäftigt sich mit Sprache und Kognition – zwei Forschungsgegenständen, die mir ans Herz gewachsen sind und die für mich immer spannend und wohl auch ein wenig unergründlich bleiben werden. Ich habe in meinen Projekten einen Zugang zu diesen Gegenständen gewählt, der die Situations- und Erfahrungsgebundenheit des Denkens und Wissens betont. Ausgangspunkt ist die Referenzialität von Sprache und die Notwendigkeit, verschiedene Informationen zu integrieren, um den Bezug zwischen Sprache, Kognition, Körper und Welt herzustellen. Meine Perspektive darauf ist durch Theorien künstlicher neuronaler Netze (Konnektionismus) und körpergebundener Kognition (Embodied Cognition) geprägt. Ziel war es, ganz im Sinne der Kognitionswissenschaft durch die Verbindung theoretischer, empirischer und synthetischer Ansätze und Methoden dem Phänomen menschlichen Sprachverstehens ein wenig näher zu kommen.

Der konnektionistischen Forschung wird mitunter unterstellt, mehr Kunst als Wissenschaft zu sein, bei der der Forscher nur lange genug modellieren muss, um einen geeigneten Klassifikator für das darzustellende Problem zu finden. In der Realität hat es meiner Erfahrung nach eher etwas von biologischer Trial-and-error-Forschung – mit vielen durchwachten Nächten und heißgelaufenen Computern. Und am Ende machen die Netzwerke doch nicht so ganz das, was man sich vorgestellt und erwartet hat. Aber gerade deshalb ist konnektionistisches Forschen spannend. Modelle sind eben keine implementierten Lösungen, sondern kleine künstliche Organismen, die der Untersuchung bedürfen. Und gerade das Ausloten der Möglichkeiten, der Grenzen und Eigenschaften der Modelle machen ihren Reiz und Nutzen für die Kognitionswissenschaft aus. Nach fünf Jahren Forschung – mitunter anstrengend, vom ‚Ph.D.-Blues‘ durchzogen und auch mal monoton, überwiegend aber spannend, lehrreich, spaßig, von vielen lieben Kollegen und Freunden begleitet, durch Lehre, psychologische Experimente und Konferenzen gewürzt – kann ich in dieser Arbeit meine Ergebnisse präsentieren. Vielleicht bringen sie uns der Antwort, wie Sprache, Kognition und Welt zusammenfinden, ein kleines Stück näher.

In meiner Zeit an der Universität hatte ich viele Begleiter, die alle auf ihre Weise ein Stück zur Entstehung meiner Arbeit sowie zu meiner persönlichen Entwicklung beigetragen haben. Mein herzlicher Dank gilt zuerst einmal meinem Doktorvater Lars Konieczny sowie meinen Kollegen vom Psycholinguistischen Forschungslabor der Universität Freiburg. Ein solches Team unter einer solchen Leitung bringt sicherlich die besten Voraussetzungen für eine erfolgreiche Promotion mit: Ständiger reger Austausch, kritische Diskussion, regelmäßige Zusammenarbeit und gegenseitige Unterstützung und nicht zuletzt eine tolle Atmosphäre unter Freunden. Die engste Zusammenarbeit verbindet mich mit Lars Konieczny, Daniel Müller, Sascha Wolfer, Peter Baumann, Sarah Schwarzkopf und Benjamin Udiljak. Ebenso schätzen gelernt habe ich meine Kollegen Wibke Hachmann, Anne Karina Feldmeth, Florian Niefind und Nina Ospelt. Euch allen vielen Dank für die gemeinsame Zeit! Hervorzuheben sind Daniel Müller und Sascha Wolfer, meine ständigen

---

Ansprechpartner bei Streitfragen zur Statistik, meine Diskussionspartner und Ratgeber für Psycholinguistik und Modellierung, und nicht zuletzt die tapferen Korrekturleser meiner Arbeit. Vielen Dank an Euch beide und viel Glück, Durchhaltevermögen und Erfolg bei der Fertigstellung Eurer eigenen Dissertation. Danke auch an Benjamin Udiljak und Florian Niefind, die meine Promotionsthemen interessant genug fanden, um daraus ihre Abschlussarbeiten in Linguistik und Informatik abzuleiten. Mein herzlicher Dank gebührt auch meinen weiteren Korrekturlesern Franziska Weldle, die die komplette Arbeit durchgeackert hat, sowie Benjamin Udiljak und Stefan Becker. Vielen Dank auch an Jochen Schnekenburger, der mir Druck und Bindung der Arbeit vermittelt hat. Nur von Luft und Wissenschaft kann man leider nicht überleben. Deshalb hing meine Promotion nicht zu gering von finanzieller Unterstützung ab. Zum Glück haben die Geber des Stipendiums der Landesgraduiertenförderung Baden-Württemberg mein Projekt als förderungswürdig beurteilt und mich für knapp drei Jahre finanziert. Vielen Dank hierfür. Alle Übergangszeiträume wurden durch mein zweites ‚Stipendium‘ überbrückt – an dieser Stelle meine tiefe Dankbarkeit an meine Eltern, die es mir ermöglicht haben, immer meinen Wünschen und Interessen nachgehen zu können. Es gibt noch viele weitere, denen ich zu danken habe – meinen Freunden, Geschwistern, Helfern und Wegbegleitern. Wer sich in der obigen Aufzählung nicht findet, dem sei versichert, dass es keine böswillige Absicht war. Abschließend, aber an erster Stelle, danke ich meiner Frau Franziska – vielen Dank für Deine Unterstützung, Deine Zeit, Deinen Trost und Deine Geduld.

Bevor ich mich und den Leser in die Tiefen der konnektionistischen Sprachverarbeitungsforschung stürze, möchte ich noch drei formale Anmerkungen zur Arbeit machen. Die erste betrifft die Sprecherperspektive: Ich habe durchgehend die Bezeichnung *wir* (anstatt *ich*) verwendet. Dies dient zum einen dazu, etwas mehr kritische Distanz zur eigenen Arbeit zu wahren. Zum anderen führt es vor Augen, dass ich mit meinen Projekten nie alleine dastand und meine Kollegen im Psycholinguistischen Forschungslabor stets an den Forschungsprojekten beteiligt waren und mir bei meiner Arbeit unterstützend zur Seite standen. Die dargestellten Ansichten und Schlußfolgerungen stellen aber ausschließlich die Meinung des Autors dar. Die zweite Anmerkung betrifft die Genderfrage: Ich bin zwar einerseits der Meinung, dass sprachlicher Geschlechtergerechtigkeit unterstützt werden sollte. Andererseits möchte ich den umfangreichen und thematisch mitunter komplexen Text nicht unnötig weiter verkomplizieren. Folglich werde ich aus Gründen der einfacheren Lesbarkeit in dieser Arbeit bei der Bezeichnung gemischter Gruppen einheitlich die alltagssprachliche männliche oder in gegebenem Fall eine neutrale Wortform wählen. Diese Wahl der Wortform dient nur der grammatikalischen Vereinfachung und ist unter allen Umständen als geschlechtsneutral zu verstehen. Die dritte Anmerkung ist eher technischer Natur: Um zu vermeiden, dieser ohnehin schon umfangreichen Arbeit nochmals über hundert Seiten Experimental- und Simulationsmaterialien hinzuzufügen, habe ich auf die Einbindung eines Anhangs verzichtet. Dieser Anhang existiert jedoch. Sollte ein Leser Interesse an der Einsicht in diese Materialien haben, bitte ich um eine Nachricht per E-Mail an [helmut.weldle@googlemail.com](mailto:helmut.weldle@googlemail.com). Ich werde ihm dann umgehend ein PDF mit dem Anhang zusenden. Und jetzt wünsche ich viel Vergnügen beim Lesen.

Frankfurt am Main, den 16.06.2011

Helmut Weldle

# 1. Einleitung

*Es gibt eine Theorie, die besagt, wenn jemals irgendwer genau herausfindet, wozu das Universum da ist und warum es da ist, dann verschwindet es auf der Stelle und wird durch etwas noch Bizarrereres und Unbegreiflicheres ersetzt. – Es gibt eine andere Theorie, nach der das schon passiert ist.*

*Douglas Adams, Das Restaurant am Ende des Universums*

Wenn wir Sprache verwenden, verweisen wir damit normalerweise auf Sachverhalte, Gegenstände oder Ereignisse in der Welt sowie innersprachlich auf andere sprachliche Elemente. Wir verwenden Sprache referenziell. Und wir tun dies mit einer solchen Selbstverständlichkeit, dass wir zumeist nicht bemerken, dass der Erwerb und die Verarbeitung von referenziellen sprachlichen Bezügen und von Bedeutung sehr komplexe Leistungen sind. Erst bei näherer Betrachtung wird deutlich, wie erstaunlich es ist, dass wir die damit verbundenen Aufgaben überhaupt lösen können. In dieser Arbeit beschäftigen wir uns mit den Problemen des Erwerbs und der Verarbeitung von Referenzialität und Bedeutung, und entwickeln Erklärungsansätze dafür, wie und mit welchen kognitiven Leistungen und Fähigkeiten wir sie bewältigen.

## 1.1. Sprache, Referenz, Kognition

In welcher Beziehung steht die Sprache zur Welt, die sie abbildet? Und wie können wir diese Beziehung und ihre Eigenschaften erklären? Während direkt beobachtbare formalstrukturelle Aspekte von Sprache relativ gut erfasst werden können, ist dies bei der Bedeutung von Sprache ungleich problematischer. Denn die Bedeutung eines Wortes oder Satzes steckt nicht im Untersuchungsgegenstand selbst, sondern in dessen Relation zu Dingen und Gegebenheiten in einer möglichen Welt. Für die Beschreibung und Erklärung von Bedeutung wurden sehr unterschiedliche Ansätze entwickelt, etwa die Referenztheorie, die Ideationstheorie, die Gebrauchstheorie und die Theorie der Wahrheitsbedingungen. Ein klassischer Zugang wird etwa mit dem triadischen Zeichenmodell vermittelt, das auf ähnliche Weise bereits in der Antike unter anderem von Aristoteles definiert und von verschiedenen Philosophen über die Jahrhunderte auf ähnliche Weise bestimmt wurde. Bekannt wurde das Semiotische Dreieck in der Sprachwissenschaft vor allem durch Ogden und Richards (1923).<sup>1</sup> Mit dem dreiseitigen Bedeutungsmodell hält das Subjekt als inter-

---

<sup>1</sup>Für die Semiotik der Neuzeit wurde das Semiotische Dreieck von C. S. Peirce eingeführt (vgl. z.B. Burch, 2010). Durch Ogden und Richards (1923) wurde das Modell dem nur zweiseitigen Modell von F. de Saussure (vgl. de Saussure, 1967) entgegengestellt. Zur Diskussion verschiedener Modelle sprachlicher Interaktion und semiotischer Grundlagen der Kommunikation vgl. Auer (1999).

pretierende Instanz Einzug in die Konzeption von Bedeutung. Es verdeutlicht, dass es keine direkte Beziehung zwischen dem sprachlichen Zeichen und den Objekten/Situationen in der Welt gibt: Das sprachliche Zeichen symbolisiert einen Sachverhalt in der Welt, diese referenzielle Beziehung muss aber durch ein interpretierendes Individuum hergestellt werden. Die Beziehung von Zeichen zur symbolisierten Welt ist also nur indirekter Natur. Das sprachliche Zeichen evoziert ein assoziiertes mentales Konzept, das auf den entsprechenden realen Sachverhalt in der Welt verweist. Was allerdings der Kern der Bedeutung ist, welchem Bestandteil des Modells sie ursächlich zugeschrieben werden soll und welcher Natur sie ist, dies wird in verschiedenen theoretischen Ansätzen sehr unterschiedlich beantwortet.

Was ist Bedeutung? In dieser Arbeit werden wir uns dieser Frage aus sprachpsychologischer und kognitionswissenschaftlicher Perspektive zuwenden. Ausgangspunkt ist die beobachtbare Fähigkeit des Menschen, im Normalfall und unter den meisten Bedingungen recht gut und erfolgreich zu verstehen, was sprachliche Äußerungen bedeuten. In verschiedenen sprachphilosophischen und linguistischen Ansätzen ist genau das die Grundlage, um zu klären, was Bedeutung ist und wodurch die Bedeutung eines Wortes oder eines Satzes determiniert ist. Von diesem Ausgangspunkt können wir fragen: Was geschieht mental, wenn ein Mensch Sprache versteht, d.h. welche kognitiven Strukturen und Mechanismen liegen dieser Verstehensleistung zugrunde? Wie erwirbt der Mensch diese Strukturen und Mechanismen? Wie analysiert er die Beziehung zwischen sprachlicher Äußerung und korrespondierendem Weltausschnitt? Wir werden uns diesen Fragen aus einer spezifischen kognitionswissenschaftlichen Perspektive nähern. Im Zentrum stehen dabei konnektionistische Modelle sowie deren Relation zu Theorien der Embodied Cognition und der Emergenz von Sprache und Kognition. Komplexe Kognition entwickelt sich in diesem Ansatz aus dem Verhalten großer Netzwerke einfacher, aber vielfach interagierender Verarbeitungseinheiten. In diesem Rahmen werden wir Modelle des Verstehens referenzieller Bezüge bei der Satzverarbeitung sowie des Erwerbs dieser Fähigkeit entwickeln. Diese Modelle werden wir zu empirischen Befunden über referenzielle Aspekte der Sprachverarbeitung sowie zu aktuellen Theorien des Verstehens in Beziehung setzen.

Wann wird der Erwerb und die Verarbeitung von Referenzialität als Untersuchungsgegenstand interessant? Zunächst einmal immer dann, wenn der referenzielle Verweis bei der Interpretation relevant wird. Dies ist zum einen beim Erwerb der Sprache der Fall, bei dem wir die korrekte Abbildung lexikalischer und grammatischer Einheiten zur Klasse ihrer möglichen Referenten etablieren müssen. Zum anderen ist es bei der situierten Sprachverarbeitung der Fall, bei der wir den zur Sprache korrespondierenden Weltausschnitt identifizieren müssen, zum Beispiel um sprachliche Aussagen zu verifizieren, um Fragen zu beantworten, oder um fehlende Bestandteile der Aussage und non-verbale Bestandteile durch Bezug auf den Weltausschnitt zu vervollständigen. Und zum dritten ist es situationsunabhängig und rein auf Sprache bezogen dann der Fall, wenn mit einem sprachlichen Ausdruck auf einen anderen Ausdruck verwiesen wird und die Interpretation des ersten vom koreferenziellen Bezug mit dem zweiten abhängig ist. Möglicherweise ist Referenzialität sogar immer relevant, wenn wir die Hypothese zugrunde legen, dass Sprachverstehen die Konstruktion eines zur Aussage korrespondierenden mentalen Situationsmodells ist, und das Erfassen der Referenz dem Verstehen der Äußerung entspricht.

Des Weiteren wird Referenzialität als Untersuchungsgegenstand interessant, wenn die einfachste Lösung einer Eins-zu-eins-Beziehung zwischen Referent und sprachlichem Aus-

druck nicht unmittelbar möglich ist. Dies kann sehr viele verschiedene Gründe haben. Ein grundlegendes Problem sind zum Beispiel Uneindeutigkeiten aufgrund verschwimmender kategorialer Grenzen, die in verschiedenen Ansätzen semantischer Theoriebildung thematisiert werden, so etwa in der Konzeption der Familienähnlichkeit, der Prototypentheorie und der sich daraus entwickelnden Prototypensemantik (vgl. Kleiber, 1993; Mangasser-Wahl, 2000; Schwarz & Chur, 2004). Das Problem liegt hier in der Unschärfe der Zuordnung von Instanzen zu einem übergeordneten kategorialen Begriff. Ein weiteres zentrales Problem ist, dass wir bei der Verarbeitung von Sprache auf verschiedensten Ebenen mit Ambiguitäten zurechtkommen müssen.<sup>2</sup>

In dieser Arbeit beschäftigen wir uns mit der Auflösung referenzieller Uneindeutigkeit in Erwerb und Verarbeitung von Sprache. Speziell untersuchen wir zwei Bereiche:

- Die Auflösung anaphorischer Referenzen. Anaphern – etwa die Pronomen *er*, *sie*, *sich* oder *ihn* – sind Verweise auf früher im Text erscheinende Argumente. Um die anaphorische Referenz aufzulösen, muss der korrekte Antezedens im sprachlichen Kontext identifiziert werden. Nur in Beziehung auf die Referenz, die durch diesen Antezedens zu Aspekten oder Objekten in einer Situation hergestellt wird, kann die aktuelle Referenz der Anapher festgestellt werden. Zunächst einmal sind Anaphern also semantisch unterspezifizierte Elemente – Variablen, die eine größere Klasse möglicher Referenten binden können. Ihr tatsächlicher Koreferent kann nur innerhalb des Äußerungskontexts bestimmt werden – und in manchen Fällen nicht einmal innerhalb von diesem endgültig. Unser Forschungsgegenstand werden pronominale Referenzen innerhalb von Sätzen sein, die anhand struktureller Beziehungen aufgelöst oder eingeschränkt werden können.
- Die Herstellung referenzieller Strukturen während der Sprachentwicklung aus der Kookkurrenz von sprachlichen Äußerungen (im Normalfall Sätze oder Phrasen) und Gegebenheiten in der Welt (Objekte, Interaktionen von Objekten, Zustände und Ereignisse in der Welt in ihren multimodalen Erscheinungsformen). Ein zentrales Problem hierbei ist die referenzielle Indeterminiertheit (Quine, 1960). Sie entsteht durch die potenzielle Viele-zu-viele-Korrespondenz: Ein Wort der jeweiligen Äußerung kann sich auf eine potenziell infinite Menge von Referenten und Aspekten der aktuellen Situation beziehen. Dadurch, dass wir in Sätzen eine ganze Menge Wörter haben, vervielfacht sich dieses Problem, denn wir müssen zugleich eine ganze Auswahl sprachlicher Symbole und deren jeweiligen Anteil an der referenzierten Situation identifizieren (vgl. Davis, 2002). Und nicht nur die einzelnen Wörter haben eine Beziehung zu Aspekten der Situation, sondern zudem auch die Relationen zwischen verschiedenen Wörtern und die syntaktische Struktur über den Wörtern.

---

<sup>2</sup>Mehrdeutigkeiten treten auf praktisch allen sprachlichen Ebenen auf. Sie finden sich etwa (a) auf phonologischer Ebene durch Homonyme, (b) auf lexikalischer Ebene durch Polysemie und Homonymie, (c) in der Wortbildung, (d) auf Ebene der Spezifizierung durch Skopusmarkierung, etwa durch Quantifizierer, und (e) auf strukturell-syntaktischer Ebene, etwa durch Anbindungsambiguitäten bei verschiedenen Möglichkeiten zur Bildung der Satzstruktur. Eine Übersicht bieten zum Beispiel Hemforth und Koniczny (2002). Eine ganze Reihe weiterer Ursachen für Uneindeutigkeiten sind dem Bereich der Pragmatik zuzuordnen. Bei diesen Phänomenen ergeben sich im Äußerungskontext Interpretationen, die stark vom semantischen Inhalt an der Oberfläche abweichen können, seien es Phraseologismen, Metaphern, Ironie oder verschiedene Formen indirekter Sprechakte.

Menschen zeigen bei der Sprachverarbeitung eine ganze Reihe an Eigenschaften. Unter anderem scheint der Spracherwerb recht unproblematisch, leicht und ohne besondere Erfordernisse oder explizite Anleitung vonstatten zu gehen. Einmal erworben, ist unser Sprachverarbeitungsvermögen sehr robust und schnell. Wir haben im alltäglichen Gebrauch kaum Schwierigkeiten, verrauschten und unvollständigen Input sinnvoll und unmittelbar zu interpretieren. Selbst tatsächliche Fehler im semantischen oder syntaktischen Gebrauch stellen uns nicht vor unlösbare Probleme. Ambige Strukturen werden zumeist nicht einmal als solche bemerkt, sondern direkt aufgrund von semantischen, pragmatischen und diskursiven Kontextinformationen aufgelöst. Allerdings funktioniert unsere Sprachverarbeitung nicht immer gleich gut: Bestimmte komplexe Strukturen und Ambiguitäten stellen uns unter bestimmten Bedingungen vor meßbar größere Schwierigkeiten als andere. Genau diese Phänomene unterschiedlicher Komplexität und Ambiguität stehen im Mittelpunkt der psycholinguistischen Forschung, denn gerade Phänomene, die zu erhöhtem Aufwand oder zum Misslingen der Verarbeitung führen, bieten eine Chance, die kognitiven Mechanismen der Sprachverarbeitung zu identifizieren.

In dieser Arbeit stellen wir Modelle vor, die die Eigenschaften der menschlichen Sprachverarbeitung modellieren. Im Zentrum stehen dabei die Erwerbs- und Verarbeitungseigenschaften für die oben genannten Gegenstände. Zudem berichten wir Ergebnisse aus empirischen Studien, die einige unserer Annahmen weitergehend untersuchen. Der Zugang zu diesen Gegenständen soll aus einer konnektionistischen Perspektive auf Sprache und Kognition erfolgen. Aus dieser werden Bezüge zu weiteren sich ergänzenden Konzepten hergestellt, in denen Kognition und Sprache als emergente Strukturen aus der Interaktion des kognitiven Systems und der physischen und sozialen Umwelt erklärt werden. Diese Interaktionsaspekte rücken auch in den Ansätzen der Embodied Cognition und deren spezifischen Sichtweisen auf die kognitiven Prozesse beim Sprachverstehen in den Vordergrund. Bezogen auf das Sprachverstehen wird in diesen die besondere Rolle der Eigenschaften des Körpers und unserer Interaktionserfahrung mit der Umwelt hervorgehoben. Diese Konzepte und Ansätze werden wir in der Diskussion einer konnektionistischen Perspektive auf die Verarbeitung von Referenzialität und den Erwerb des Sprachverstehens näher beleuchten.

## 1.2. Ausgangspunkte und Projekte der Arbeit

Untersuchungsgegenstand dieser Arbeit ist der Erwerb und die Verarbeitung von sprachlicher Referenzialität und darauf aufbauend das Verstehen sprachlicher Äußerungen auf Satzebene. Unser Ausgangspunkt für die kognitionswissenschaftliche Beschreibung, Analyse und Erklärung der hierbei angenommenen Prozesse, Repräsentationen und Verarbeitungsprinzipien ist der Konnektionismus in der Tradition lernender Modelle mit parallel-verteilter Verarbeitung (im Sinne des PDP-Ansatzes, Rumelhart & McClelland, 1986b; McClelland & Rumelhart, 1986). In dieser Forschungstradition wurden Konzeptionen von Kognition, Sprache und Bedeutung entwickelt, die stark von den damals vorherrschenden symbolisch-funktionalistischen Ansichten abwichen. Auf diesen Konzeptionen aufbauend skizzieren wir eine konnektionistische Perspektive für spezielle Aspekte der natürlichen Sprachverarbeitung und entwickeln Erklärungsansätze für das dabei vom Menschen gezeigte Verhalten. Im Zentrum stehen dabei die oben dargestellten Gegenstände der Auflösung anaphorischer

Referenzen und der Etablierung referenzieller Korrespondenz bei situational-sprachlicher Vieldeutigkeit im Kontext syntaktischer Verarbeitung.

Für unsere Ansätze verwenden wir als zentrale Architektur die sogenannten Simple Recurrent Networks (SRNs, Elman, 1990), die sich in den vergangenen 20 Jahren als klassisches konnektionistisches Modell der Verarbeitung sequenzieller Strukturen etabliert haben. Für diese Netzwerke können zwei häufige Verwendungsweisen unterschieden werden: die Verwendung mit Vorhersage-Aufgabe oder mit Abbildungs-Aufgabe. Die erstgenannte ist die klassische Aufgabe für diesen Netzwerk-Typ. Das Netzwerk muss hierbei die Vorhersage des jeweils nächsten Wortes einer Sequenz leisten und darüber die Antizipation der im sequenziellen Kontext möglichen Folgewörter erlernen. Dieser Leistung liegt eine rekurrent eingeschleifte Gedächtniskomponente zugrunde, die dem Netz die benötigte Sensitivität für den sequenziellen Kontext und die Erfassung hierarchischer Abhängigkeiten ermöglicht. Diese Form der Modellierung erfasst aber keine Referenzialität: Semantische Strukturen bildet das Netzwerk nur als Gruppierungen sequenzieller Ähnlichkeit ab, im Sinne intensionaler rein sprachlicher Strukturen auf Basis der Kollokationen von Wörtern. Damit bleiben viele Aspekte der sprachlichen Verarbeitung und der hierbei beobachtbaren Performanz außer Acht – unter anderem alle Aspekte, die den Bezug zwischen sprachlicher Form und nicht-sprachlicher referenzieller Bedeutung herstellen, und damit für die Modellierung unseres Untersuchungsgegenstands von zentraler Bedeutung sind. Die zweitgenannte Aufgabe bildet sequenzielle Muster auf gänzlich andere Repräsentationsstrukturen ab, die in irgendeiner Form die Bedeutung der jeweiligen Sequenz darbieten. Auf diese Weise wird in sehr unterschiedlichen Modellierungsansätzen sprachliche Referenzialität und das Verstehen von Sätzen simuliert.

SRNs legen eine Reihe von Eigenschaften und Prinzipien der Verarbeitung nahe, die für die Perspektive unseres Ansatzes relevant werden, unter anderem die Fähigkeit zur Antizipation von Folgezuständen, die gleichzeitige Berücksichtigung verschiedener Informationen, Beschränkungen und des Kontexts, die Sensitivität für die statistischen Eigenschaften des Inputs und nicht zuletzt die Fähigkeit zur Integration verschiedener Informationsstrukturen und Aufgaben (vgl. z.B. Altmann & Mirkovic, 2009). Auf diesen Eigenschaften begründend haben wir zwei Modelle entwickelt, die in verschiedener Weise die wechselseitige Abbildung und Integration verschiedener Aufgaben und Informationsstrukturen leisten und dabei Aspekte der Referenzialität von Sprache abbilden.

Das erste Modell *AnaSRN* erweitert die klassische SRN-Architektur, um die Simulation der referenziellen Verarbeitung bei der Anaphernauflösung zu ermöglichen. Wir untersuchen auf dieser Grundlage die Auflösung von Relativ-, Personal- und Reflexivpronomen. Dies ist eine interessante Herausforderung für die konnektionistische Modellierung, da anaphorische Referenzen sowohl lexikalisch als auch syntaktisch komplex sind (vgl. R. Frank, Mathis & Badecker, 2005): Als Form der Variablenbindung haben Anaphern einen variablen Referenzgegenstand, der nur durch den Kontext identifiziert werden kann. Diese kontextabhängige Bindung des Referenzgegenstands erfolgt allerdings nach strukturellen Constraints, die von anderen syntaktischen Dependenzien unabhängig sind. Neben der allgemeinen Leistung untersuchen wir spezifische Fehlleistungen des Modells und Interferenz-Effekte auf die Anaphernauflösung. Dabei werden auch kritische Aspekte wie die Unfähigkeit zur Abbildung der Identitätsfunktion thematisiert (vgl. G. F. Marcus, 1998; Elman, 1998).

Mit besonderem Gewicht auf der Beurteilung des Einflusses des lokalen syntaktischen Kontexts lassen sich aus dem Modell *AnaSRN* konkrete empirische Vorhersagen zur Verarbeitung von Anaphern ableiten, die in Beziehung zu aktuellen psycholinguistischen Debatten stehen. Thematisiert wird zum einen die Verarbeitung lokaler syntaktischer Kohärenzen (im Weiteren: LSK), für die ein Einfluss des lokalen syntaktischen Kontexts auf die Interpretation von Sätzen festgestellt wurde (Tabor, Galantucci & Richardson, 2004). Die Interpretation der lokalen Sequenz kann dabei temporär im Widerspruch zur global konsistenten Interpretation stehen (Konieczny, Müller, Hachmann, Schwarzkopf & Wolfer, 2009). Zum anderen lassen sich Eigenschaften für die Online-Verarbeitung von Anaphern vorhersagen, die den zeitlichen Verlauf der Auflösung, mögliche Verletzungen der Annahme globaler strukturell bedingter Konsistenz, potenzielle Interferenz-Quellen und die Erwartung unterschiedlicher Verarbeitungseigenschaften für verschiedene anaphorische Formen betreffen (vgl. u.a. Badecker & Straub, 2002; Sturt, 2003; Runner, Sussman & Tanenhaus, 2006; Kaiser, Runner, Sussman & Tanenhaus, 2009). Für die Konzeption einer größeren empirischen Studie haben wir Kernthemen und Annahmen aktueller Debatten herangezogen und zusammen mit den Vorhersagen aus der Modellierung mit *AnaSRN* eine Reihe von Hypothesen abgeleitet. Diese Hypothesen haben wir anhand der Methoden Lesezeitmessung im Self-Paced-Reading-Paradigma, Eyetracking im Visual-World-Paradigma und Acceptability Rating im Projekt *AnaLoCo* untersucht. Die Befunde zeigen unter anderem, dass sich LSK auf die Auflösung anaphorischer Referenzen auswirken. Daraus ergeben sich Konsequenzen für das Phänomen der LSK-Effekte sowie für die allgemeinen Eigenschaften anaphorischer Verarbeitung: Zum einen beeinflussen LSK die Ebene der referenziellen Interpretation und weisen damit auf grundlegende übergreifende Verarbeitungsprinzipien hin, die auf lokaler Ebene operieren. Zum anderen scheint bei der Verarbeitung von Anaphern bereits sehr früh eine initiale Auswahl möglicher Referenten zu erfolgen, bei der zudem die globale Konsistenz struktureller Constraints verletzt werden kann.

Das zweite Modell *SynSem* erweitert SRNs auf eine Weise, die die Verarbeitung und Integration zweier unterschiedlicher Routen ermöglicht: Eine sprachliche Route, auf der sequenziell Wort für Wort Sätze verarbeitet werden, und eine visuelle Route, auf der Objekte erkannt, lokalisiert und in eine räumliche Relation gestellt werden. Dazu verbinden wir SRNs mit einer Dual-Route-Architektur (vgl. Cangelosi, 2005) und erweitern damit frühere konnektionistische Modelle des Erwerbs und der situationalen Verankerung auf der lexikalischen Ebene (vgl. z.B. Plunkett, Sinha, Møller & Strandsby, 1992).<sup>3</sup> Die Aufgabe des Netzwerks ist es zunächst, eine Mikrosprache situational verankert in einer visuellen Mikrowelt zu erwerben. Die Mikrosprache besteht aus kurzen beschreibenden Mehrwort-Äußerungen, die visuellen Situationen bestehen aus einfachen Punktmustern auf einer zweidimensionalen pseudo-visuellen Retina, deren Anordnung partiell zur jeweiligen parallel dargebotenen sprachlichen Äußerung korrespondiert. Diese Reduzierung der verwendeten Stimuli soll sicherstellen, dass dem Modell keine strukturellen und konzeptuellen Informationen gegeben werden, die über das hinausgehen würden, was Kinder beim natürlichen Spracherwerb erhalten: umweltbezogene Mehrwort-Äußerungen mit uneindeutiger Satz-Bedeutungs-Korrespondenz. Obwohl es ein sehr primitives und in vielfacher Hinsicht

---

<sup>3</sup>Die dabei entstehende rekurrente Dual-Route-Architektur weist Bezüge zum theoretischen Ansatz von Altmann und Mirkovic (2009) auf.

vereinfachendes Modell ist, genügt dieser Ansatz, um grundlegende Fragen des situierten Spracherwerbs sowie der kontextbezogenen und der kontextfreien Sprachverarbeitung zu modellieren.

Die Konzeption von Bedeutung, die mit diesem Ansatz verfolgt wird, zeigt viele Bezüge zu theoretischen Ansätzen der Embodied Cognition.<sup>4</sup> Aktuelle Ansätze in diesem Forschungsfeld lehnen die Annahme von Kognition als abstrakt-amodalem konzeptuellem System ab. Sie stellen dieser Sichtweise die Annahme entgegen, dass sich Kognition in der Interaktion von System, Körper und Umwelt entwickelt. Kognitive Prozesse und Repräsentationen sind direkt in der Wahrnehmung der Umwelt und in den motorischen Interaktionsmöglichkeiten mit der Umwelt verankert (Barsalou, 1999; Glenberg, 1997). Kognitive Verarbeitung entspricht dabei der kontrollierten mentalen Simulation von Derivaten früherer Erfahrungen. Sprachverstehen wird auf dieser Grundlage als mentale Simulation eines zur Äußerung korrespondierenden Situationsmodells verstanden (Simulationstheorie der Bedeutung, vgl. z.B. Bergen, 2007; Zwaan, 2004). Die Konstruktion dieses Situationsmodells erfolgt durch die Kontrolle über sprachliche Hinweisreize, mit denen die Erfahrungsderivate indiziert sind (vgl. Glenberg & Robertson, 2000). Wir werden diskutieren, inwiefern Konnektionismus und Embodied Cognition kompatible Ansätze sind, die sich gegenseitig ergänzen können.

Unser konnektionistischer Ansatz im Projekt *SynSem* kann als Modellierung des Embodied Language Processing verstanden werden: Sprache wird in modalen analogen Repräsentationen der wahrgenommenen Umwelt verankert, Sprachverstehen erfolgt als Simulation von schematischen Wahrnehmungsderivaten. Wir untersuchen dabei die Performanz des Modells unter verschiedenen Bedingungen: Welche Leistung ergibt sich, wenn es nur sprachlichen Input erhält (Sprachverstehen ohne situationalen Kontext), wenn es nur visuellen Input erhält (Sprachproduktion bei der Situationsbeschreibung) und wenn es beide Informationen erhält (situiertes Sprachverstehen)? Wir zeigen dabei, wie sich aus der Integration verschiedener Informationen und Aufgaben als emergentes Produkt der Interaktion weitere kognitive Verarbeitungsleistungen entwickeln – etwa Antizipation und Inferenzen, modalitätsübergreifendes Priming oder Steuerung der visuellen Aufmerksamkeit. Zugleich implementiert das Modell eine Lösung des Abbildungsproblems von Sätzen auf komplexe Szenen. Es funktioniert dabei anhand einer systematischen szenenübergreifenden Kombinatorik (vgl. statistical cross-situational learning, Yu & Smith, 2006; Gleitman, 1990), bei der anhand der Überschneidungen zwischen mehreren Satz-Situations-Abbildungen Wahrscheinlichkeiten für Eins-zu-eins-Korrespondenzen zwischen sprachlicher Form und (nicht-sprachlicher) Bedeutung etabliert werden.

### 1.3. Übersicht zur Gliederung der Arbeit

Die Arbeit ist in drei größere Blöcke gegliedert: **Teil I** beschäftigt sich in drei Kapiteln mit dem Themengebiet Konnektionismus und Sprache. Wir skizzieren eine konnektionistische Perspektive auf Bedeutung, semantische Repräsentation und Verstehensprozesse, und be-

---

<sup>4</sup>Viele Termini aus der Embodied Cognition lassen sich ungemein schwer ins Deutsche übersetzen. Deshalb werden wir in dieser Arbeit die englischen Begriffe *Embodiment*, *Embodied Cognition*, *Embodied Language Processing* und *Language Grounding* beibehalten.

sprechen Modelle der Satzverarbeitung und des Sprachverstehens. **Teil II** stellt nach einem einleitenden Kapitel über empirische Befunde zu Satzverarbeitung und Anaphernaufflösung in zwei Kapiteln die Projekte *AnaSRN* und *AnaLoCo* vor. In diesen Projekten untersuchen wir mit konnektionistischen und psycholinguistischen Methoden die Beziehung zwischen dem Phänomen lokaler syntaktischer Kohärenzen und der Online-Verarbeitung von Anaphern. **Teil III** untersucht die Zusammenhänge von Embodied Cognition, Language Grounding, Konnektionismus und Sprachverstehen. In zwei Kapiteln stellen wir die Ansätze der Embodied Cognition vor und diskutieren ihr Verhältnis zu Sprachverstehen und Konnektionismus. In einem dritten Kapitel stellen wir das Projekt *SynSem* vor, in dem ein Ansatz zur Modellierung von situationaler Sprachverankerung und Emergenz von Prozessen des Sprachverstehens entwickelt wird. Die folgenden Abschnitte geben einen Überblick über die Inhalte der einzelnen Kapitel.

Teil I: In **Kapitel 2** präsentieren wir zunächst die Grundlagen des Konnektionismus. Wir besprechen zentrale Fragen zu Repräsentation und Bedeutung in künstlichen neuronalen Netzen, unter anderem zur Parallelverarbeitung und zur verteilten Repräsentation konzeptuellen Wissens. Ein kurzer Überblick erläutert konnektionistische Ansätze zu Erwerb, Verarbeitung, Repräsentation und Entwicklung von Sprache und die Idee einer konnektionistischen Psycholinguistik (vgl. Christiansen & Chater, 2001a). In **Kapitel 3** vertiefen wir die Simulation sequenzieller Verarbeitung mit SRNs (Elman, 1990). Dabei besprechen wir Leistungspotenzial und Eigenschaften, interne Repräsentation von lexikalischem und grammatischem Wissen, Erwerb und Verarbeitung komplexer syntaktischer Strukturen, sowie Effekte von Distanz, Frequenz, Regularität und lokaler Kohärenz (u.a. Elman, 1991; Rohde & Plaut, 1999; Christiansen & Chater, 1999b; MacDonald & Christiansen, 2002; Konieczny, Ruh & Müller, 2009). In Exkursen erläutern wir Fehlermaße für die Evaluation der Netzwerkperformanz und Methoden der Analyse der internen Repräsentation, die wir in den später besprochenen Modellierungsprojekten heranziehen. In **Kapitel 4** stellen wir zentrale Aspekte der Modellierung von Sprachverstehen auf Satzebene und ihre Umsetzung in älteren und aktuellen Ansätzen vor (u.a. Rohde, 2002; Mayberry, 2003; Mayberry, Crocker & Knoeferle, 2009; S. L. Frank, Haselager & van Rooij, 2009). Neben den Zielen der verschiedenen Ansätze stehen die Realisierung der Abbildung von sprachlicher Form auf (nicht-sprachliche) Bedeutung, die zugrundegelegte Repräsentation von Bedeutung und die Kontrolle der Konstruktion der Satzbedeutung im Verstehensprozess im Vordergrund.

Teil II: In **Kapitel 5** erläutern wir psycholinguistische Grundlagen zu den Themen Syntaxverarbeitung und Anaphernaufflösung, die wir für die folgenden projektbezogenen Kapitel benötigen. Dazu berichten wir zunächst psycholinguistische Studien zum Phänomen LSK und diskutieren die Ergebnisse der aktuellen Forschung zum Thema (u.a. Tabor et al., 2004; Konieczny, Müller, Hachmann et al., 2009). Anschließend wenden wir uns der Online-Verarbeitung anaphorischer Referenzen zu (u.a. Badecker & Straub, 2002; Sturt, 2003; Runner et al., 2006; Kaiser et al., 2009). In **Kapitel 6** schlagen wir einen Ansatz zur Modellierung von Referenzzuweisung und Anaphernaufflösung anhand eines erweiterten SRNs vor, den wir aufgrund der Kritik an einer früheren Studie (R. Frank et al., 2005) entwickelt haben. Anschließend stellen wir eine Serie von Simulationen mit dieser Architektur vor (vgl. auch Weldle, Baumann, Müller, Wolfer & Konieczny, 2009). Besonderes Interesse lag bei den Untersuchungen auf der Erfassung struktureller Constraints bei der Auflösung anaphorischer Referenzen, dem Einfluss struktureller Interferenzen

und den Eigenschaften der dabei gebildeten Generalisierungen. Anhand der Simulationen werden wir empirisch überprüfbare Hypothesen zur Anaphernauflösung ableiten, aber auch Einschränkungen von SRNs (vgl. z.B. G. F. Marcus, 1998) diskutieren. In **Kapitel 7** berichten wir eine Serie empirischer Untersuchungen zur Anaphernverarbeitung bei Personal- und Reflexivpronomen sowie zum Einfluss von LSK hierauf. Fragestellungen und Hypothesen leiten wir aus Befunden früherer psycholinguistischer Studien und aus den Ergebnissen unserer konnektionistischen Simulationen ab. Die Studie umfasst eine Serie mehrerer Experimente (vgl. Weldle, Konieczny, Wolfer, Baumann & Müller, 2010; Konieczny, Weldle, Wolfer, Müller & Baumann, 2010) – eine Lesezeitmessung im Self-Paced-Reading-Paradigma, zwei Eyetracking-Experimente im Visual-World-Paradigma und eine ergänzende Rating-Studie. Unsere Befunde weisen darauf hin, dass die Auflösung anaphorischer Referenzen zumindest bei der initialen Verarbeitung sensitiv für den lokalen Kontext innerhalb eines Satzes ist und die globale Konsistenz der koreferenziellen Bindung hierdurch verletzt werden kann.

Teil III: In **Kapitel 8** führen wir in die theoretischen Ansätze der Embodied Cognition und ihre Perspektive auf das Verstehen von Sprache ein. Wir skizzieren Theorien der Konzeptualisierung von Gedächtnis, Verarbeitung und Kategorisierung (v.a. Lakoff & Johnson, 1980; Glenberg, 1997; Barsalou, 1999), und erläutern die daraus hervorgehenden Modelle zum Sprachverstehen (u.a. Glenberg & Robertson, 2000; Zwaan, 2004). Anschließend berichten wir empirische Studien zum Embodied Language Processing und gehen dabei auf Erklärungsansätze für die Verarbeitung von situationaler Perspektivität und anaphorischer Referenz ein, die auf der Konzeptualisierung von Verstehen als mentaler Simulation der beschriebenen Situation basieren (Schwarzkopf, Müller, Weldle & Konieczny, 2008; Brunyé, Ditman, Mahoney, Augustyn & Taylor, 2009; Konieczny, Haser et al., 2010; Konieczny, Weldle, Haser et al., 2010). In **Kapitel 9** beleuchten wir die Beziehungen zwischen Konnektionismus und Embodied Cognition, ihre Kompatibilität und mögliche Synergien einer gemeinsamen Perspektive. Dabei besprechen wir Modellierungsansätze, die eine Implementierung von Ideen des Language Grounding und der Embodied Cognition anstreben (u.a. Plunkett et al., 1992; Cangelosi, 2005; S. L. Frank et al., 2009). Abschließend formulieren wir Kriterien einer konsistenten Modellierung von Sprachverarbeitung in der Konzeption der Embodied Cognition. **Kapitel 10** berichtet eine Serie konnektionistischer Simulationen zur situationalen Verankerung von Spracherwerb und -verarbeitung und zur Emergenz von Eigenschaften und Fähigkeiten des Sprachverstehens (vgl. Weldle, Konieczny, Müller, Wolfer & Baumann, 2009a, 2010a, 2010b). Die Architektur der Modelle basiert auf einer Kombination von Dual-Route-Netzwerken (Cangelosi, 2005) und rekurrenten Netzwerken (Elman, 1990). Ziel ist die Modellierung von Sprachverankerung auf syntaktisch-sequenzieller Ebene. Darüber hinaus wird mit einem pseudo-visuellen Input ein strikt nicht-propositionaler Ansatz verfolgt. Die Modellierungen zeigen, wie Korrespondenzen zwischen sprachlicher Form und (nicht-sprachlicher) Bedeutung etabliert werden und wie sich dabei als Nebenprodukt integrativer Verarbeitung verschiedener Informationen und Aufgaben Verarbeitungseigenschaften wie Aufmerksamkeitslenkung, Inferenzen und Unterspezifizierung entwickeln.

Im abschließenden **Kapitel 11** fassen wir Ziele und Ergebnisse der Arbeit zusammen und erläutern unsere Schlussfolgerungen zur entwickelten konnektionistischen Perspektive auf Grundfragen referenzieller Verarbeitung von Sprache.



## **Teil I.**

# **Konnektionismus und Sprache**



## 2. Konnektionismus

Konnektionismus ist eine kognitionswissenschaftliche Forschungsrichtung, die Aspekte menschlicher Kognition mithilfe künstlicher neuronaler Netzwerke zu modellieren und über diese Modelle zu erklären sucht. Das Forschungsprogramm verfolgt einen prozessorientierten Ansatz und wird mitunter als Gegenentwurf zur repräsentationsorientierten Forschung in der Kognitionswissenschaft charakterisiert (vgl. Pospeschill, 2004), denn es lehnt die Annahme syntaktisch-algorithmischer Manipulationen als Operationen auf diskreten Symbolen ab. In diesem Kapitel wird der Konnektionismus als kognitive Theorie in seiner Perspektive auf Verarbeitung, Repräsentation und Sprache diskutiert. Nach einer kurzen Einführung zu Grundlagen der kognitiven Modellierung werden wir auf die Konzeption parallel-verteilter Verarbeitung und subsymbolischer Repräsentation eingehen. Dabei besprechen wir auch, wie im Konnektionismus Aspekte der Bedeutung – etwa Kategorisierungsprozesse, konzeptuelles Wissen, Kompositionalität und Systematizität – erfasst werden.

In Vorbereitung auf die beiden folgenden Kapitel stellen wir Konnektionismus als Perspektive auf Sprache und als Ansatz zur Modellierung von Sprachverarbeitung vor. Im Zusammenhang mit dem Gegenstand Sprache hat diese Forschungsrichtung neue Debatten auf einer großen Bandbreite ausgelöst – betreffend Repräsentation (Regeln und diskrete Symbole/Strukturen vs. Assoziationen und verteiltes Wissen), Verarbeitung (Produktionen und Regeln vs. Constraint-Satisfaction-Prozesse) und Erwerb (Nature-Nurture-Debatte und induktiv-assoziatives Lernen) bis zur Evolution von Sprache (Emergenz, Sprache als dynamisches System). In Bezug auf die daraus abgeleiteten Eigenschaften der Sprachverarbeitung und ihre empirische Überprüfung wird mitunter von einer konnektionistischen Psycholinguistik gesprochen (Christiansen & Chater, 2001a, 2001b).

### 2.1. Kognitive Modellierung und Konnektionismus

Kognitive Modellierung ist eine zentrale, wenn nicht sogar eine definierende Methode der Kognitionswissenschaft. Sie lässt sich kurzgefasst als Versuch charakterisieren, kognitionswissenschaftliche Modelle als lauffähige Computerprogramme zu formulieren<sup>1</sup> und damit kognitive Prozesse zu simulieren, die die zu untersuchenden Verhaltensaspekte oder Theorien abbilden (vgl. u.a. Strube, 1998; Simon & Wallach, 1999). Die Simulationen werden zu Implementierungen der kognitionswissenschaftlichen Theorien (Cangelosi &

---

<sup>1</sup>Dabei darf die kognitive Modellierung nicht allgemein mit computergestützter Modellierung gleichgesetzt werden: Es geht in der Zielsetzung nicht nur darum, eine bestimmte Leistungsfähigkeit zu erzeugen (im Sinne einer effizienten Lösung für ein gegebenes Problem), sondern Verhalten und interne Prozesse des zu untersuchenden Systems nachzubilden (im Sinne einer Nachbildung des Lösungswegs, nach dem das zu modellierende Vorbild vermutlich operiert).

Parisi, 2002). Es geht darum, anhand des Modells und anhand wechselseitiger Evaluationen zwischen Theorie, Modell und Empirie die Repräsentationen und Mechanismen zu identifizieren, die unseren kognitiven Fähigkeiten zugrunde liegen.

### 2.1.1. Gütekriterien und Ziele der Modellierung

Simon und Wallach (1999) formulieren Gütekriterien für kognitive Modelle hinsichtlich ihrer Korrespondenz zu den empirisch festgestellten Phänomenen der menschlichen Kognition. Anhand dieser Kriterien können die Modelle evaluiert werden. Dabei werden folgende Korrespondenzen gefordert: (a) Korrespondenz des Produkts – kann das Modell ein bestimmtes Verhalten erzeugen und dem empirischen Vorbild entsprechende Lösungen finden? (b) Korrespondenz der Zwischenschritte – werden die gleichen Lösungsstrategien und -abfolgen gewählt? (c) Korrespondenz der Zeit – zeigen sich vergleichbare Latenzen und zeitliche Verläufe beim Lösen des Problems? (d) Korrespondenz des Fehlers – produziert das Modell während des Lösens der Aufgabe gleich häufig gleichartige Fehler wie das empirische Vorbild? (e) Korrespondenz der Kontextabhängigkeit – beeinflussen Aspekte des Kontextes die Performanz des Modells in gleicher Weise? (f) Korrespondenz des Lernens – zeigt sich der gleiche Entwicklungsverlauf beim Lernen einer Problemlösung oder beim Erwerb einer Fähigkeit?

Einerseits ist dieser Katalog sehr umfassend, und viele Modelle nehmen gezwungenermaßen Abstraktionen des Gegenstands vor, die nicht die Evaluation aller Korrespondenzkriterien ermöglichen. Zudem ist die Feststellung der Korrespondenz kaum objektiv zu erfassen und hängt davon ab, welche Eigenschaften des Modells von Forscher zum Vergleich herangezogen werden. Andererseits ist der Katalog natürlich bei weitem nicht vollständig und sehr allgemein gehalten. Gerade für das Themenfeld Sprache – der Spracherwerb, die Entwicklung von Sprache über die Zeit und die Evolution der Sprachfähigkeit – müssten einige erforderliche Korrespondenzen hinzugefügt bzw. spezifiziert werden, deren Untersuchung relevanz für den Forschungsgegenstand hat. Um nur zwei Kandidaten zu nennen: (g) Korrespondenz der Lernbarkeit der Prozesse und Strukturen (als Erweiterung von Punkt f). Die Frage dahinter ist: Kann das Modell die gleichen Strukturen lernen, die das empirische Vorbild lernen kann, sowie die gleichen Strukturen nicht oder nur eingeschränkt lernen, die auch dem empirischen Vorbild Schwierigkeiten bereiten oder bei diesem nicht beobachtet werden (vgl. z.B. die Studien zu sprachlichen Universalien und typologischen Merkmalen von van Everbroeck, 2003; Christiansen & Devlin, 1997). (h) Korrespondenz der Veränderlichkeit des Gegenstands. Die Frage ist hier, ob das Modell die Veränderlichkeit des Problems/Gegenstands erlaubt, und ob es dessen Entwicklung nachvollziehen kann (vgl. z.B. die Studie zur historischen Entwicklung der englischen Verbmorphologie von Hare & Elman, 1995).

Als Ziele der Modellierung können wir kurzgefasst Beschreibungsadäquatheit, Erklärungsadäquatheit und Transparenz festhalten (vgl. Pospeschill, 2004, S. 64). Es geht nicht nur darum, das äußerlich beobachtbare Verhalten abzubilden, sondern darüberhinaus sein Entstehen zu erklären und seine Wirkungszusammenhänge offenzulegen. Die kognitive Modellierung beinhaltet dabei eine Reihe von Vorteilen und Nutzen für die Theoriebildung, die den erheblichen zeitlichen und technischen Aufwand rechtfertigen, den sie mit sich bringt:

- Konkretisierung und Spezifizierung von Theorien: Um die Theorie in einem lauffähigen Modell zu formulieren, muss dieses gezwungenermaßen in allen zu modellierenden Aspekten explizit, detailliert, in sich konsistent, und vollständig gefasst sein. Es ist kein Raum für Vagheit und Auslassungen. Dies macht die Theorie besser überprüfbar bzw. empirisch falsifizierbar, unterstützt und konkretisiert die Aussagen der Theorie, und ermöglicht eine genaue Identifizierung ihrer Fehler und Schwächen.
- Beleg der Konsistenz und der internen Validität von Theorien: Ein lauffähiges Modell beweist die Konsistenz der internen Logik bzw. der internen Widerspruchsfreiheit des Modells. In den Worten von Simon und Wallach (1999, S. 2): „First of all, a running generative model provides an existence proof for the sufficiency of its structures and mechanisms for a given task, hence, a first strict falsification test for the model.“
- Erklärungsansatz theoretischer und empirischer Zusammenhänge: Bei geeigneter Korrespondenz der Performanz zu den modellierten Daten erbringen die im Modell formalisierten Mechanismen einen legitimen Erklärungsansatz für das zu erklärende Phänomen.
- Generative Theorien: Computermodelle sind darüber hinaus generativ, d.h. sie können über bisher gesehene Probleme hinaus generalisieren und für neue, noch nicht gesehene Probleme Verhalten produzieren. Dieses Verhalten entspricht der Vorhersage der Theorie für neue Fälle, die empirisch überprüft werden können. Damit stimulieren die Modelle neue empirische Forschung. Durch die empirische Überprüfung der Vorhersagen können umgekehrt die Modelle evaluiert werden – Modellierung und empirische Forschung sind also sich wechselseitig ergänzende und stimulierende Methoden.
- Erfassung komplexer Zusammenhänge: Viele empirische Phänomene und Aspekte von Theorien werden durch so viele Einflussfaktoren und Interaktionen bedingt, dass man sie gar nicht sinnvoll über verbal formulierte Theorien erfassen kann (vgl. Rohde, 2002). Hier bieten Simulationen die Möglichkeit, diese Interaktionen abzubilden und gegebenenfalls besser zu verstehen. Dabei kann die Simulation auch helfen, die zugrundeliegende Theorie zu konkretisieren und weiterzuentwickeln.
- Simulation als virtuelles Forschungslabor: Ein einmal implementiertes Modell kann immer wieder für Experimente in einer maximal kontrollierten Umgebung herangezogen werden, um den Einfluss beliebiger Manipulationen von Bedingungen und Variablen zu untersuchen (Cangelosi & Parisi, 2002). Anders als in empirischen Studien können dabei auch Bedingungen untersucht werden, die in der realen Welt nicht möglich wären, etwa kontrafaktische oder nicht kontrollierbare Bedingungen, sowie Phänomene, die sich über einen sehr langen Zeitraum erstrecken. Dies macht Computersimulationen zu einem sehr praktischen Werkzeug, zum Beispiel für die Erforschung der Evolution der Sprachfähigkeit.

### 2.1.2. Einordnung konnektionistischer Modellierung

Inwiefern treffen diese Eigenschaften auf konnektionistische Modelle zu? Und auf welche spezifische Weise können sie dazu beitragen, beobachtete Phänomene adäquat und

transparent zu beschreiben und zu erklären oder Theorien zu konkretisieren? Hier ist anzumerken, dass die Rolle des Konnektionismus innerhalb der Kognitionswissenschaft und seine Möglichkeiten für die Entwicklung erklärungsadäquater Theorien mitunter auch von Vertretern dieses Forschungsbereichs sehr kritisch betrachtet wird. McCloskey (1991) beispielsweise betrachtet konnektionistische Modelle als Methodik zur Entwicklung von Theorien, verweigert ihnen aber den Status einer realisierten Theorie bzw. einer Simulation derselben. Allein die Architektur und die technischen Spezifika der Simulation machen ihm zufolge noch keine Theorie aus, denn es fehlt die Identifikation der relevanten und entscheidenden Anteile daraus – die Beschreibung des Netzwerks liefert keine inhaltliche Erklärung der Prozesse. Zudem zeigen sich relevante Unterschiede zwischen konnektionistischer Modellierung und Modellierung im klassischen Sinne: Zum einen kann die Theorie vage und unspezifisch sein, zum anderen können auch erfolgreich abgeschlossene Simulationen aufgrund ihrer Komplexität mitunter nicht auf theoretische Positionen bezogen werden. Konnektionistische Netzwerke sind aus dieser Perspektive eher als Modelle im Sinne von Untersuchungsobjekten zu verstehen, an denen Theorien überprüft werden können.

An dieser Stelle müssen wir den Gegenstand näher eingrenzen, denn konnektionistische Modelle erscheinen in vielen – auch inkompatiblen – Unterarten, auf die die in den folgenden Abschnitten besprochenen Charakterisierungen nur teilweise oder sogar überhaupt nicht zutreffen. Wenn im Weiteren von konnektionistischen Modellen gesprochen wird, so sind in der Regel eliminative, nicht-strukturierte, lernende Netzwerke im Sinne des Parallel Distributed Processing (PDP, Rumelhart & McClelland, 1986b; McClelland & Rumelhart, 1986) gemeint. Von diesen lernenden Modellen unterscheiden sich zum Beispiel Modelle mit fest verdrahteten Gewichten.<sup>2</sup> Im Unterschied zu Modellen mit verteilter Repräsentation gibt es lokalistische Modelle, die auf der Konzeption sogenannter Großmutterzellen aufbauen (zu einer aktuellen Debatte hierzu vgl. Bowers, 2009). Des Weiteren klammern wir Modelle des implementierenden und strukturierten Konnektionismus aus, mit denen versucht wird, im Sinne von Symbolsystemen operationale Regeln und diskretes Wissen zu implementieren. Die von uns untersuchten konnektionistische Modelle gehören prinzipiell zu den lernenden Modellen und sind dem PDP-Ansatz zuzurechnen. Als lernende Modelle sind sie nicht zu voller Funktionalität vorsezifiziert, sondern entwickeln ihre Verarbeitungseigenschaften in Auseinandersetzung mit den Trainings-Stimuli, d.h. in der Interaktion mit ihrer Umwelt. Erst nach der Trainingsphase kann ihre Performanz mit empirischen Daten und Leistungsprofilen verglichen werden.

Was spezifizieren wir in konnektionistischen Modellierungen hinsichtlich Repräsentationen und Mechanismen? Wie übertragen wir die theoretischen Konzepte, die wir modellieren wollen, in ein konnektionistisches Modell? Und wie können wir die Korrespondenz der Netzwerkperformanz zu empirischen Daten oder Theorieaspekten überprüfen? In der kognitionswissenschaftlichen Anwendung wird versucht, die Performanz natürlicher kognitiver Systeme über konnektionistische Netze zu modellieren und in Analogie zu deren Eigenschaften tiefere Einblicke in die Funktionsweise des kognitiven Systems zu erhalten.

---

<sup>2</sup>Ein Beispiel für ein konnektionistisches Modell mit spezifiziert-fixierten Gewichten wäre etwa das bekannte Interactive Activation and Competition Network (auch bekannt als Jets-and-Sharks-Modell, vgl. McClelland & Rumelhart, 1981). Auch dieses ist dem PDP-Ansatz zuzurechnen, allerdings ist es eben kein lernendes Modell.

Anders als bei klassisch-symbolischen Modellen<sup>3</sup> werden keine konkreten internen Repräsentationen und keine expliziten Zwischenschritte oder Teilziele vorgegeben. Es werden auch keine Mechanismen im Sinne von expliziten Operationen oder Regeln implementiert. Vielmehr wird die Architektur anhand von Neuronenschichten und Verknüpfungsmustern zwischen diesen entwickelt. Die allgemeine Funktionsweise wird anhand übergreifender Aktivierungsfunktionen und Gewichtsanpassungsmechanismen spezifiziert. Zugleich werden durch Bestimmung des Inputs und mitunter des Outputs die Aufgaben des Netzwerks und die externe Datenstruktur, mit der das Netzwerk arbeiten muss, vorgegeben.

Das spezifische Verhalten von künstlichen neuronalen Netzwerken resultiert aus den generellen Eigenschaften des gesamten Modells (vgl. u.a. McClelland & Rumelhart, 1986; Plunkett & Elman, 1997; McLeod, Plunkett & Rolls, 1998; Pospeschill, 2004). Aufgrund inhärenter Eigenschaften entwickeln die Modelle interessante Verhaltensweisen und generieren selbstorganisierend mitunter unerwartete Lösungen für komplexe Probleme. Diese Eigenschaften, auf die wir in den folgenden Abschnitten näher eingehen werden, machen sie zu geeigneten und interessanten Modellen zur Simulation verschiedener Aspekte der Sprachverarbeitung auf großer Bandbreite: von Sprachproduktion und -rezeption auf vielen linguistischen Ebenen (von phonetischer Diskrimination bis zu inkrementellem Satzverstehen), über den Erwerb von Sprache bis hin zu Untersuchungen von Typologie, Sprachwandel und Sprachgenese.

Im Unterschied zu symbolischen Modellen gibt es in konnektionistischen Modellen keine explizite Trennung von Verarbeitung und Repräsentation. Die Annahme syntaktisch-algorithmischer Regeln, die auf diskreten Wissensseinheiten operieren, wird aufgegeben. An deren Stelle tritt die Weiterleitung von Aktivierungspotentialen, die Aktivierungsmuster ansprechen, die über das gesamte System verteilt sind (sogenannte subsymbolische Repräsentationen, vgl. Kap. 2.2.2). Diese alternative Sicht auf kognitive Strukturen und Prozesse bietet eine neue Perspektive auf kognitive Phänomene, die von klassischen Erklärungsansätzen abweicht. Beispielsweise legen die konnektionistischen Modelle der bereichsspezifischen Verarbeitung allgemeine bereichsunspezifische Mechanismen zugrunde – zur Erklärung der Sprachfähigkeit legen sie folglich generelle, übergreifende kognitive Fähigkeiten nahe.

## 2.2. Konnektionismus und Kognition

Mit dem konnektionistischen Ansatz wird versucht, auf der Grundlage weniger Annahmen und Prinzipien alle benötigten Konzepte zur wissenschaftlichen Beschreibung und Erklärung kognitiver Phänomene zu formulieren: Repräsentation, Konzeptualisierung und

---

<sup>3</sup>Im Allgemeinen werden konnektionistische Modelle immer wieder pauschalisierend Modellen aus der Tradition der klassischen Künstliche-Intelligenz-Forschung gegenübergestellt. Unter diese auch als GOFAI (good old fashioned artificial intelligence, z.B. M. Anderson, 2003) bezeichneten Tradition werden vereinfachend klassische (unifizierende) Architekturen etwa auf Grundlage von Produktionensystemen wie zum Beispiel CAPS, ACT-R, SOAR oder PSI subsummiert. Kennzeichen sind u.a. die explizite konzeptuelle Unterscheidung von Wissensstrukturen und Verarbeitungsprozessen sowie die Annahmen, dass Repräsentation auf diskreten abstrakten Symbolen beruht und Verarbeitung syntaktisch-algorithmischen Operationen auf diesen Symbolen entspricht. Wir werden in diesem Zusammenhang im Folgenden immer von klassisch-symbolischen Modellen reden.

Strukturierung, Verarbeitung, kognitive Ressourcen und Constraints (zu Ansätzen einer konnektionistischen Kognitionspsychologie vgl. u.a. McClelland & Rumelhart, 1986; Elman et al., 1996; Pospeschill, 2004). Dabei wird in prinzipieller Abgrenzung von klassischen symbol-orientierten Konzepten der Kognition – etwa im Sinne einer *Language of Thought* (Fodor, 1976; Fodor & Pylyshyn, 1988) oder eines *Physical Symbol System* (Newell, 1980, 1990) – auch von subsymbolischer Verarbeitung (Smolensky, 1988) und von parallel-verteilter Verarbeitung (Parallel Distributed Processing, Rumelhart & McClelland, 1986b; McClelland & Rumelhart, 1986) gesprochen.<sup>4</sup> Im Kern bleibt der Versuch, kognitive Prozesse anhand einfacher assoziativer Mechanismen und allgemeiner bereichsunspezifischer Verarbeitungs- und Repräsentationsprinzipien zu erklären und damit vorhandene komplexere Erklärungsansätze zu überdenken.

### 2.2.1. Parallele Verarbeitung

Konnektionismus ist zunächst grundlegend als ein Berechnungsparadigma zu verstehen. Er gibt verschiedene Möglichkeiten vor, Inputmuster in Outputmuster zu transformieren, die vor allem als assoziativ und statistisch (mit einem Spektrum von allgemein linear bis nicht-linear dynamisch) charakterisiert werden können. In dieser Hinsicht führt der Konnektionismus als Theorie kognitiver Systeme das zentrale Paradigma der Kognitionswissenschaft fort, nach dem kognitive Prozesse als Berechnungsprozesse zu verstehen sind. Verarbeitung erfolgt in diesem Rahmen als Propagierung von Aktivierungsmustern durch das Netzwerk. Zentrale Eigenschaft ist, dass ein konnektionistisches System keine komplexen zentralen Berechnungskomponenten zur Verfügung stellt, sondern aus einer Vielzahl kleiner, sehr einfach operierender Einheiten besteht, die hochgradig miteinander verknüpft sind und eine massive parallele Verarbeitung leisten. Deshalb spricht man auch synonym zum Begriff Konnektionismus von Parallel Distributed Processing (PDP, Rumelhart & McClelland, 1986b; McClelland & Rumelhart, 1986). Damit wird in einer gewisse Analogizität die biologische Realität menschlicher und tierischer Kognition nachgebildet: Das Vorbild des Konnektionismus ist die biologische Implementierung von Kognition in einem zentralen Nervensystem, das seine Leistungsfähigkeit einer großen Anzahl einfacher aber komplex interagierender Nervenzellen verdankt. In diesem Zusammenhang wird auch von der Theorie künstlicher neuronaler Netze (KNN) gesprochen, wobei aber stark von natürlichen neuronalen Netzen abstrahiert wird.

Als Einstieg in die im weiteren Verlauf verwendete Terminologie und als Zusammenfassung einiger zentraler Konzepte besprechen wir kurz formale Aspekte der Aktivierung und

---

<sup>4</sup>Dabei stellt sich natürlich die Frage, ob sich sowohl der symbolische als auch der konnektionistische Ansatz in diesem Maße pauschalisieren lassen, denn weder folgen alle symbolischen Theorien starren Einschränkungen ihrer Eigenschaften und Funktionsprinzipien, noch lassen sich alle konnektionistischen Ansätze der subsymbolischen Verarbeitung subsumieren. Die beiden Ansätze lassen sich eher als Extreme auf einem Kontinuum theoretischer Ansätze beschreiben (vgl. Hilario, 1997; Holyoak & Hummel, 2000). So ist der pauschalisiert dargestellte Symbolismus wohl eher als Mittel zur Kontrastierung und Konkretisierung konnektionistischer Konzepte zu verstehen.

des Lernens am Beispiel einfacher mehrschichtiger Netzwerke.<sup>5</sup> Anschließend führen wir einige zentrale Verarbeitungseigenschaften auf, die diese Modelle charakterisieren.

### 2.2.1.1. Zentrale Konzepte und Grundbegriffe

Abbildung 2.1 zeigt ein Feedforward-Netzwerk (im Weiteren: FFN) mit drei Neuronenschichten. Die Grundeinheiten des Modells sind formale Neuronen, die im Folgenden als Units bezeichnet werden. Die Units sind gruppiert in Neuronenschichten, die wir im Folgenden als Layer bezeichnen. Die Units sind miteinander verbunden. Die Verbindungen werden im Folgenden als Gewichte bezeichnet. Über die Gewichte wird zwischen verbundenen Units Aktivierung weitergeleitet, dabei wird die Aktivierung durch die Stärke des jeweiligen Gewichts transformiert. Der Aktivierungsfluss erfolgt nicht kontinuierlich, sondern in diskreten Zeitschritten. Im FFN wird Aktivierung nur in eine Richtung – in Abbildung 2.1 von unten nach oben<sup>6</sup> – und immer nur von Layer zu Layer weitergeleitet, nicht aber innerhalb der Layer. Aktivierung fließt also vom Input-Layer zum zwischengeschalteten Layer (im Folgenden: Hidden-Layer) und von dort weiter zum Output-Layer. Soweit nicht anders angegeben, sind die Layer vollständig verknüpft, d.h. jede Unit eines Layers ist mit jeder Unit des jeweils nächsten Layers durch ein Gewicht verbunden.

Nebenstehender Kreis in Abbildung 2.1 zeigt die Grundfunktionen einer Unit  $i$ . Diese empfängt von  $n$  sendenden Units  $j$  Aktivierung  $a$  über die Gewichte  $w_{ij}$ . Die empfangende Unit  $i$  summiert die Produkte der Aktivierung  $a$  der Unit  $j$  und des Gewichts zwischen sendender und empfangender Unit  $w_{ij}$ . Diese Summe der eingehenden Aktivierung ist der Netzwerk-Input  $net_i$ . Die Unit verarbeitet diesen Netzwerk-Input  $net_i$ , indem sie ihn anhand einer Aktivierungsfunktion  $F(net_i)$  transformiert (Gleichung 2.1). Das Resultat der Funktion über dem Netzwerk-Input bestimmt den Output der Unit  $i$ . Gängige Aktivierungsfunktionen sind binäre, lineare und logistische (sigmoide) Funktionen.<sup>7</sup> Der Output einer Unit, der über ausgehende Gewichte an weitere Units im nächsten Layer weitergegeben wird, entspricht in der Regel dem Wert der Aktivierung der Unit (Identitätsfunktion).<sup>8</sup>

<sup>5</sup>Es geht dabei nicht um Vollständigkeit, sondern darum, grundlegende Prozesse zu benennen und Begriffe zu bestimmen, auf die wir im späteren Verlauf zurückgreifen. Vertiefungen der Grundlagen finden sich in der einschlägigen Fachliteratur. Eine gute Einführung in den Konnektionismus bieten zum Beispiel McLeod et al. (1998), Standardwerke zum Konnektionismus sind Rumelhart und McClelland (1986b); McClelland und Rumelhart (1986); Elman et al. (1996), eine deutschsprachige Einführung in den Konnektionismus hinsichtlich seiner Eigenschaften als Modell menschlicher Kognition bietet Pospeschill (2004).

<sup>6</sup>Die Abbildung der Netzwerke mit Aktivierungsrichtung von unten nach oben ist die Konvention für weitere Abbildungen in dieser Arbeit.

<sup>7</sup>Bei einer linearen Funktion erfolgt eine lineare Transformation des Netzwerk-Inputs. Wird für diese Funktion kein Limit verwendet, kann das Resultat für den Output der Unit beliebig große und kleine Werte annehmen. Bei einer binären Funktion wird über einen Schwellenwert ein binäres Resultat bestimmt. Je nachdem, ob der Netzwerk-Input diese Schwelle überschreitet oder nicht, nimmt der Wert einen von zwei festen Werten an (z.B. 0 oder 1). Eine häufig verwendete logistische Funktion ist die Sigmoid-Funktion. Hier ist das Resultat ebenfalls nach oben und unten begrenzt, nimmt dabei aber kontinuierlich beliebige Zwischenzustände in einem sigmoiden Verlauf an. Für Extremwerte nähert sich das Resultat der Funktion asymptotisch den Grenzwerten an.

<sup>8</sup>Der Output kann aber auch weiter verändert werden, zum Beispiel über eine Normalisierungsfunktion (Ausgabe aller Units eines Layers summiert zu 1, etwa über eine Softmax-Funktion), eine Winner-takes-all-Funktion (nur die am stärksten aktivierte Unit leitet Aktivierung weiter, alle anderen nicht) oder eine

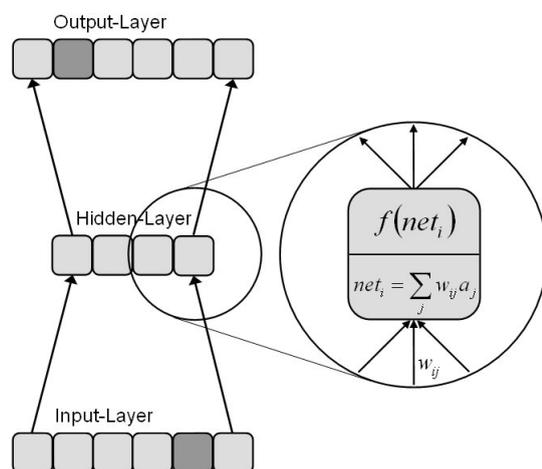


Abbildung 2.1.: Architektur eines einfachen mehrschichtigen Netzwerks (FFN). Aktivierung fließt vom Input-Layer über ein Hidden-Layer zum Output-Layer. Dort liegt ein Target-Vektor an, mit dem die Output-Aktivierung verglichen wird. Der Fehler wird über den Backpropagation-Algorithmus über den Hidden-Layer zurück zum Input-Layer geleitet. Anhand des zugewiesenen Fehlers werden die Gewichte angepasst. Nebenstehender Kreis zeigt ein formales Neuron, für das die Grundfunktionen vermerkt sind.

$$f(net_i) = f\left(\sum_j w_{ij} \cdot a_j\right) \quad (2.1)$$

Die Leistung, die das Netzwerk zunächst erbringen soll, ist die Abbildung einer Menge von Aktivierungsmustern des Input-Layers auf eine Menge von Aktivierungsmustern des Output-Layers. Das Netzwerk erfüllt damit die Aufgabe der Musterassoziation. Die Aktivierungsmuster können auch als Vektoren von Aktivierungsstärken der linear angeordneten Units des jeweiligen Layers aufgefasst werden. In künstlichen neuronalen Netzwerken erfolgt Lernen in der Regel über eine Anpassung der Gewichte.<sup>9</sup> Durch diese Anpassung wird die Weiterleitung der Aktivierung so verändert, dass die daraus resultierende Aktivierung des Output-Layers einem Zielzustand entspricht. Im Falle des überwachten Lernens ist dieser Zielzustand fest als Target-Vektor vorgegeben.<sup>10</sup>

Noise-Funktion (der Output wird mit einem Rauschen versehen, indem die Aktivierung innerhalb eines vorgegebenen Rahmens zufällig verändert wird).

<sup>9</sup>Alternativ bzw. zusätzlich kann der Schwellenwert der Aktivierungsfunktion der einzelnen Units angepasst werden. Etwas ähnliches wird in Netzwerken mit ausschließlicher Anpassung der Gewichte durch ein gängiges Verfahren erreicht: Alle veränderlichen Layer (i.d.R. alle Layer außer dem Input-Layer und speziellen Copy-Layern) werden mit einer Bias-Unit verknüpft, die immer Aktivierung sendet. Über die veränderliche Stärke der Gewichte von der Bias-Unit wird eine Anpassung des Schwellenwerts der einzelnen Units simuliert.

<sup>10</sup>Neben dem überwachten Lernen werden alternative Assoziationsaufgaben und Lernformen verwendet. Eine hinsichtlich des Vorgehens praktisch identische Form ist das auto-assoziative Lernen. Hier erfolgt das Lernen selbstüberwacht: Das Netzwerk lernt, die Input-Vektoren auf jeweils identische Output-Vektoren abzubilden. Beim unüberwachten Lernen hingegen wird kein definierter Zielzustand des Output-Vektors vorgegeben. Stattdessen wird mittels Hebbscher (verstärkendes Lernen) oder kompetitiver Lernprozeduren eine Kategorisierung der Input-Vektoren angestrebt, für die das Netzwerk über verschiedene Output-Aktivierungsmuster eine eigene Lösung entwickelt.

Zum Lernen des vorgegebenen Target-Vektors müssen die Gewichte angepasst werden. Dazu wird der Output der Units des Output-Layers mit den vorgegebenen Werten des Target-Vektors verglichen, indem die Differenz des Outputs jeder Output-Unit  $a_{out}$  zum korrespondierenden Wert des Target-Vektors  $t_{out}$  berechnet wird (Gleichung 2.2). Diese Differenz ist der Fehler  $\delta_i$  des Outputs der Unit  $i$ . Je nachdem, ob der tatsächliche Output im Vergleich zum gewünschten zu gering oder zu groß ist, werden auf Grundlage dieses Fehlers die Stärken der Gewichte zu dieser Unit verringert oder erhöht. Hierfür bestehen verschiedene alternative Lern-Algorithmen, die in Abhängigkeit von der Netzwerk-Topologie, der Lernform (überwacht vs. unüberwacht) und dem gewünschten Verhalten ausgewählt werden.

$$\delta_i = (t_{out} - a_{out}) \quad (2.2)$$

Ein grundlegender Lern-Algorithmus hierfür ist die Deltaregel. In dieser wird die Gewichtsanzpassung  $\Delta w_{ij}$  für ein Gewicht zwischen einer Unit  $j$  und einer Unit  $i$  als Produkt des Fehlers des Outputs der Output-Unit  $\delta_i$ , der Output-Aktivierung der sendenden Unit  $a_j$  und einer vorgegebenen Lernrate  $\epsilon$  berechnet (Gleichung 2.3). Über die Lernrate  $\epsilon$  wird die Stärke der Anpassung der Gewichte je Lernschritt reguliert und damit die Geschwindigkeit und die Präzision des Lernens gesteuert. Dabei sind in Abhängigkeit der Größe des Netzwerks, der Eigenschaften der zu lernenden Stimuli und der zu leistenden Aufgaben verschiedene Lernraten sinnvoll. Durch die Aufnahme des Faktors  $a_j$  werden nur Gewichte von Units angepasst, die tatsächlich Aktivierung zugeführt haben. Je nachdem, ob  $\delta_i$  einen negativen (bei zu starker Output-Aktivierung) oder einen positiven Wert (bei zu geringer Output-Aktivierung) annimmt, wird der vorhandene Wert des Gewichts  $w_{ij}$  um den Wert von  $\Delta w_{ij}$  erhöht oder verringert. Die Deltaregel implementiert ein Gradientenabstiegsverfahren<sup>11</sup>, über das der Fehler des Outputs auf ein Minimum reduziert wird.

$$\Delta w_{ij} = \epsilon \cdot \delta_i \cdot a_j \quad (2.3)$$

Ein zentrales Problem der frühen konnektionistischen Forschung war, dass das zu dieser Zeit technisch realisierbare Perzeptron (einschichtiges Netzwerk ohne Hidden-Layer mit nur einer Schicht anpassbarer Gewichte, vgl. Rosenblatt, 1958) nur linear separierbare Probleme lösen konnte, während nicht linear separierbare Probleme – wie etwa das klassische XOR-Problem – nicht bewältigt werden konnten (vgl. Minsky & Papert, 1972). Diese Erkenntnis führte zum zeitweiligen Niedergang des Konnektionismus, bis etwa zu Beginn der 1980er Jahre neue Verfahren entwickelt wurden, mit denen dieses Problem bewältigt werden konnte. Die Lösung brachte der Lern-Algorithmus Backpropagation (Werbos, 1974; Rumelhart, Hinton & Williams, 1986), der das Training mehrschichtiger Netzwerke ermöglichte. Anhand der Hidden-Layer können die Netzwerke die nicht linear separierbaren Probleme der Transformierung von Input-Aktivierungsmuster zum gewünschten Output-Aktivierungsmuster über eine Reorganisation der Repräsentation (d.h. die Entwicklung geeigneter interner Repräsentationen für Zwischenlösungen) der Aus-

<sup>11</sup>Die Deltaregel entspricht im Verfahren der Least Mean Squares nach Widrow und Hoff (1960) der Ableitung ersten Grades der Aktivierungsfunktion, und implementiert damit ein Gradientenabstiegsverfahren für die Quadratsummenabweichung für lineare Aktivierungsfunktionen.

gangszustände so verändern, dass sie anhand der entwickelten Zwischenrepräsentationen (als linear separierbare Probleme) gelöst werden können.

Die Schwierigkeit des Trainings mehrschichtiger Netzwerke liegt darin, dass der Wert zur Anpassung der Gewichte nur in Abhängigkeit zum Fehler der Aktivierung der jeweils empfangenden Unit bestimmt werden kann. Dies ist für die Gewichte zwischen Hidden- und Output-Layer möglich, nicht aber für die Gewichte zwischen Input- und Hidden-Layer, da für die Hidden-Layer keine vorbestimmten Target-Vektoren bekannt sind. Durch den Backpropagation-Algorithmus werden Schätzwerte für den Anteil der Units im Hidden-Layer am Fehler der Units im Output-Layer bestimmt. Damit kann entsprechend der Vorwärts-Propagierung der Aktivierung (Gleichung 2.1) die Rückwärts-Propagierung des Fehlers erfolgen. Hierfür wird der Fehler  $\delta_j$  der Unit  $j$  als Funktion über der Summe der Produkte der Fehler  $\delta_i$  aller Units  $i$  (die von  $j$  Aktivierung empfangen) und ihrer Gewichte  $w_{ij}$  berechnet (Gleichung 2.4). In Abhängigkeit der Stärke der Gewichte und des jeweils entstandenen Fehlers der empfangenden Units wird damit der Unit  $j$  ein Fehler zugewiesen, der den Anteil an der Verursachung des Folgefehlers der empfangenden Units abschätzt, und der die Berechnung der Anpassung der Gewichte  $\Delta w$  zwischen der versteckten Unit  $j$  und den vorausgehenden (zu Unit  $j$  Aktivierung sendenden) Units ermöglicht. Die Weiterentwicklung von Backpropagation mit dem Algorithmus Backpropagation Through Time (BPTT, Werbos, 1990; Rumelhart, Hinton & Williams, 1986) ermöglicht die Speicherung und Zusammenfassung des Fehlers für mehrere aufeinanderfolgende Zeitschritte. Damit können die Gewichte in einem Schritt für mehrere Stimuli angepasst werden. Dies erlaubt ein komputational sparsameres und hinsichtlich des Lernverlaufs gleichmäßigeres Lernen.

$$\delta_j = f\left(\sum_i w_{ij} \cdot \delta_i\right) \quad (2.4)$$

Der Backpropagation-Algorithmus implementiert als verallgemeinerte Deltaregel ein Gradientenabstiegsverfahren. Ein Nachteil dieses Verfahrens ist, dass die Lösung von Abbildungsproblemen damit zwar grundsätzlich möglich ist, zugleich aber nicht garantiert werden kann, dass die Lösung gefunden wird. Das System kann sich innerhalb lokaler Minima – nicht-optimalen Lösungen oberhalb des minimalen Fehlers – stabilisieren.

### 2.2.1.2. Zentrale Eigenschaften

Der Konnektionismus hat sich als Alternative zu klassischen symbolisch-algorithmischen Kognitionstheorien etabliert. Gründe dafür sind die Eigenschaften konnektionistischer Modelle, die in verschiedener Hinsicht die Eigenschaften menschlicher Kognition sehr gut abbilden, sowie eine Vielzahl an Simulationen zu kognitiven Kernfähigkeiten und Phänomenen menschlichen Verhaltens. Künstliche neuronale Netzwerke zeigen anhand inhärenter Eigenschaften Leistungsmerkmale, die wir auch beim Menschen beobachten können: Hierzu gehören sehr schnelle (Echtzeit-)Verarbeitung großer und komplexer Datenmengen (Planung komplexer Handlungsabläufe, visuelle Verarbeitung, etc.), robuste Verarbeitung (Toleranz gegenüber verrauschtem Input, allmähliche schrittweise Abnahme der Funktionalität bei Störungen im System), schneller Abruf von Informationen (durch inhaltsadressierte Speicherung mit beliebigem Abruf-Hinweis), kontextsensitives und da-

tenbasiertes Lernen sowie erfahrungsbasierte Generalisierung bzw. Prototypisierung der Umwelt. Der Konnektionismus eröffnet als subsymbolische Alternative neue Perspektiven auf eine Reihe von Domänen, die bislang eher durch symbolorientierte Erklärungsansätze dominiert wurden, und ermöglicht eine Neukonzeption der Erklärung kognitiver Prozesse. In diesem Zusammenhang wurden Ansätze herangezogen, die Kognition im Sinne assoziativer und induktiver Lernmodelle, dynamischer Systeme, probabilistischer und statistischer Modelle, Constraint Satisfaction und Emergenz erklären. Wir fassen kurz zentrale Verarbeitungseigenschaften zusammen (ausführliche Diskussionen finden sich u.a. in Rumelhart & McClelland, 1986b; McClelland & Rumelhart, 1986; Rumelhart, 1989; McClelland, 1989; van Gelder & Port, 1995; Elman et al., 1996; Elman, 1998; MacWhinney, 1998, 1999a; Pospeschill, 2004; MacDonald & Seidenberg, 2006):

**Parallel-verteilte Verarbeitung** Eine wesentliche Eigenschaft konnektionistischer Modelle ist, dass die Verarbeitung von Informationen nicht seriell (sukzessive über mehrere Zeitschritte hinweg) für verschiedene Aspekte der Informationen oder für verschiedene Teilaufgaben bei der Verarbeitung dieser Informationen erfolgt, sondern parallel: Mehrere Verarbeitungseinheiten sind gleichzeitig damit beschäftigt, alle Aspekte und damit verbundene Teilprozesse synchron und interaktiv zu verarbeiten. Damit implementieren konnektionistische Modelle das Prinzip der Constraint Satisfaction (vgl. Rumelhart, Smolensky, McClelland & Hinton, 1986; McClelland, St. John & Taraban, 1989; MacDonald & Seidenberg, 2006): Bei der Verarbeitung wirken verschiedenartige Constraints auf verschiedenen Ebenen und ermöglichen damit komplexe Abbildungen und die Interaktion und Integration unterschiedlicher Informationen.

**Datengetriebene (statistische) Verarbeitung** Lernen und Verarbeitung erfolgen datengetrieben, d.h. die Modelle entwickeln ihre Wissensstruktur und ihre Verarbeitungseigenschaften während des Trainings durch die Interaktion mit Beispielen. Der Lernprozess erfolgt assoziativ und induktiv (vgl. Elman, 2005), neue Wissensstrukturen werden aus den Eigenschaften des Inputs abgeleitet. Die Verarbeitung neuer Informationen wird aber durch bereits erworbene Wissensstrukturen beeinflusst. Die Modelle sind sensitiv für die distributionalen Eigenschaften des Inputs und erfassen dessen statistische Zusammenhänge über verschiedene Beispiele sowie über verschiedene Ebenen der Granularität hinweg (vgl. Seidenberg & MacDonald, 1999; Elman, 2003). In dieser Hinsicht gibt es vielfach Bezüge zwischen konnektionistischen Modellen und statistischen Klassifikationsverfahren.

**Selbstorganisation** Beim Lernen entwickelt das konnektionistische System selbstorganisiert, d.h. ohne leitende bzw. zentral gesteuerte Prozesse, Strukturen zur Verarbeitung des Inputs. In dieser Hinsicht implementieren konnektionistische Modelle auch das Prinzip der Emergenz: Das funktionierende System entwickelt sich aus der Interaktion von System und Umwelt und durch die Wirkung interagierender Constraints (vgl. Elman, 1998; MacWhinney, 1998; Ellis, 1998; Stephan, 1999). Dabei leistet es Generalisierung (hierarchische Klassifizierung des Inputs) und Prototypisierung (Gradualität und zentrale Tendenzen kategorialer Zugehörigkeit).

**Nicht-lineare dynamische Verarbeitung** Die selbstorganisierte Entwicklung von internen Strukturen zur Repräsentation und Transformation ermöglicht mehr als nur primitive Musterassoziationen – zumindest seit dem Aufkommen mehrschichtiger

rekurrenter Netzwerke. Die Modelle leisten nicht-lineare Transformationen komplexer Zusammenhänge zwischen Input- und Output-Strukturen. In diesem Sinne implementieren sie nicht-lineare dynamische Systeme (vgl. van Gelder & Port, 1995; Elman, 1995; Elman et al., 1996; van Gelder, 1998).

### 2.2.2. Verteilte Repräsentation

Weitere zentrale Eigenschaften konnektionistischer Modelle ergeben sich aus den Spezifika ihrer Repräsentation bzw. ihrer internen Struktur und der Speicherung von Wissen. Hierzu ist zunächst zwischen lokalistischer und verteilter Repräsentation zu unterscheiden. Bei lokalistischer Repräsentation wird der Aktivierung einzelner Units Bedeutungshaftigkeit zugeschrieben. Ist eine Unit aktiviert, so wird der ihr zugewiesene Bedeutungsgehalt (z.B. *Großmutter*) verarbeitet bzw. abgerufen. Der größte Teil der konnektionistischen Forschungsgemeinschaft (v.a. in der PDP-Tradition) hält lokalistische Repräsentationen aber für behavioral und neurobiologisch unplausibel<sup>12</sup> – sie werden nur als Enkodierungsformat für Input- und Output-Strukturen verwendet.

Der lokalistischen steht die verteilte Wissensspeicherung gegenüber (zum Begriff der verteilten Repräsentation vgl. z.B. van Gelder, 1992). Anstatt durch bedeutungshafte Einzelaktivierungen ist Repräsentation hier durch bedeutungshafte Aktivierungsmuster realisiert, die über eine größere Anzahl von Units verteilt sind. Nur die Komposition mehrerer Einzelaktivierungen repräsentiert konzeptionelles Wissen. Die Aktivierung einzelner Units innerhalb dieser Muster hingegen kann semantisch-konzeptionell nicht mehr sinnvoll interpretiert werden und einzelne Units sind an der Repräsentation verschiedener Bedeutungsinhalte beteiligt.<sup>13</sup> Da jede Repräsentation als Aktivierungsmuster über mehrere Units realisiert ist, gibt es keinen grundlegenden Unterschied mehr zwischen einfachen und komplexen Repräsentationen und auch keine atomaren Bedeutungseinheiten im Sinne von Symbolen. Kognitive Verarbeitung über verteilten Repräsentationen erfolgt auf subsymbolischer Ebene (vgl. Smolensky, 1988). Der verteilten Repräsentation wird ein höherer Grad biologischer Plausibilität zugesprochen, da sie wichtige Eigenschaften ermöglicht, die der Informationsverarbeitung natürlicher kognitiver Systeme zugeschrieben werden (zu einer entgegengesetzten Position vgl. Bowers, 2009). Vor allem ermöglicht sie ein robustes Verarbeitungsverhalten: Die Modelle sind – unter der Voraussetzung verteilt repräsentierten Inputs – tolerant gegenüber verrauschtem, unvollständigem oder partiell widersprüchli-

---

<sup>12</sup>Dennoch wird in der Forschung weiterhin an Theorien auf Grundlage lokalistischer Repräsentationen (sog. Großmutterzellen) festgehalten. Bowers (2009) etwa versucht, die (vielfach missverstandenen) Verarbeitungseigenschaften lokalistischer Modelle zu erklären, und erläutert in einem Review neurophysiologischer Daten, dass die Eigenschaften dieser Modelle eine wesentlich höhere Deckung zu Einzelzelleitungen aufweisen. Weitergehend stellt er dabei die behavioral-neurobiologische Plausibilität verteilter Modelle in Frage. Eine kritische Diskussion und Gegenposition zu dieser Besprechung findet sich in Plaut und McClelland (2010).

<sup>13</sup>Diese Aussage muss allerdings relativiert werden: Der Anteil einer Unit an einem Bedeutungsaspekt kann mitunter sehr wohl semantisch-konzeptionell interpretiert werden, und zum Teil kann die Aktivierung einer einzelnen Unit ausschließlich ein einziges Konzept repräsentieren. Nur ist dies nicht mehr automatisch und von Anfang an gegeben, und es ist auch nicht mehr entscheidend für die Realisierung der Repräsentation.

chem Input und zeigen eine graduelle Abnahme der Leistungsfähigkeit und Präzision. Zudem sind sie tolerant gegenüber dem Ausfall einzelner Units und Verbindungsgewichte.

Für das Substrat der Bedeutung – den eigentlichen Träger des konzeptuellen Wissens – gibt es mehrere Kandidaten (Diskussionen finden sich in Peschl, 1996; Garson, 2010). So werden etwa wie eingangs erläutert die Aktivierungszustände des Netzwerks – d.h. die Aktivierungsmuster über Units der (internen) Layer – als Repräsentationen verstanden. Mathematisch betrachtet können diese Aktivierungsverteilungen als Vektoren aufgefasst werden: Die  $n$  Units eines Layers repräsentieren die Vektorkomponenten, die Aktivierung der Units bestimmen den numerischen Wert der Vektorkomponente. Jedes Aktivierungsmuster (erfasst als Vektor) entspricht einem Repräsentationsobjekt, das im hochdimensionalen Raum ( $n$ -dimensionaler Vektorraum) verortet ist. Alle Objekte nehmen in diesem Raum Positionen ein, deren Verortung verglichen werden kann – etwa hinsichtlich ihrer Distanzen, ihrer räumlichen Gruppierung oder Orthogonalisierung. Diese Verortung ist ein weiterer Kandidat als Repräsentationsträger: Repräsentationen werden als Zustände im hochdimensionalen Raum (Repräsentations- oder Aktivierungsraum) verstanden (vgl. State Space Semantics, Churchland, 1986). Die distinktive Bedeutung der Repräsentationen ist dabei nicht in Aktivierungen oder den Aktivierungsmustern selbst verankert, sondern in den Beziehungen (Ähnlichkeiten, Distinktionen und Distanzen) zwischen Repräsentationen und Repräsentationsclustern im hochdimensionalen Raum. Diese Ähnlichkeitsstruktur im Repräsentationsraum entspricht dem Kategorisierungswissen und spiegelt die relevanten kategorialen Relationen der Umwelt wider.<sup>14</sup>

Ein weiterer Kandidat für die Repräsentation in konnektionistischen Netzwerken sind die Gewichte, über die die Aktivierungen propagiert und damit die Aktivierungsmuster erzeugt werden (vgl. z.B. Jackson & Sharkey, 1996). In diesen sind die kumulierten Erfahrungen dauerhaft als Zustände gespeichert, die diese Aktivierungsmuster erzeugen bzw. die Aktivierungsausbreitung determinieren. Veränderungen der Gewichte beim Lernen verändern die Dynamik des Systems und damit sein Wissen. Die Gewichte bestimmen sozusagen adaptiv die Transformationsfunktion für Umweltreize. Und schließlich können Repräsentationen basierend auf der Annahme repräsentationaler Zustände im hochdimensionalen Aktivierungsraum als Trajektorien durch diesen Zustandsraum verankert werden (vgl. z.B. Elman, 1991). Damit kann die Repräsentation komplexer Ereignisse erfasst werden. Dabei werden sowohl die Aktivierungsmuster als auch die Gewichte relevant: Die Aktivierungsmuster spiegeln die Übergangszustände der Trajektorie wider. Der Verlauf der Trajektorie wird aber dynamisch durch Attraktoren gebildet, die durch den Zustand der Gewichte determiniert werden.

Allerdings ist in Frage zu stellen, ob man bei konnektionistischen Netzwerken von Repräsentation im Sinne optimaler Umweltabbildung sprechen kann, oder es sich eher um eine hinreichende adaptive Verhaltensgenerierung handelt (vgl. Peschl, 1996). In diesem Zusammenhang wurde der Konnektionismus den symbolischen Theorien als antirepräsentationalistische Theorie gegenübergestellt. So verweist zum Beispiel Pospeschill (2004)

---

<sup>14</sup>Laakso und Cottrell (2000) diskutieren im Sinne der State Space Semantics (Churchland, 1986), wie in unterschiedlichen Systemen (z.B. unterschiedlich große Hidden-Layer) und auf Grundlage verteilter Repräsentationen durch die Ähnlichkeitsstruktur im Repräsentationsraum äquivalente Repräsentationsstrukturen (z.B. für die Gliederung von Farben) entstehen können (für eine kritische Gegenanalyse vgl. Calvo Garzón, 2008).

auf die Abkehr des Konnektionismus von der Repräsentationsorientierung klassischer symbolverarbeitender Theorien hin zu einer Prozessorientierung. Symbolische Repräsentationen zeichnen sich durch folgende Eigenschaften aus: Sie sind strukturiert (auf explizit systematische, dekomponierbare Weise durch Symbole und syntaktische Kombinationsregeln festgelegt), kontextfrei (Wissenseinheiten werden unabhängig vom Kontext ihres Auftretens gespeichert), intern-semantisch (bedeutungsunterscheidende Merkmale sind formal in der Struktur der Wissenseinheit verankert) und passiv (Repräsentationen werden nur verändert, wenn direkt mit ihnen operiert wird). Im Konnektionismus hingegen können Repräsentationen dahingehend charakterisiert werden, dass sie holistisch (die aktuelle Repräsentation entsteht durch Aktivierung einer distinktiven Gruppierung von Verarbeitungseinheiten), kontextgebunden (der Beitrag der einzelnen Verarbeitungseinheiten steht immer in Relation zum gesamten Beitrag), extern-semantisch (die Bedeutung entspricht der Adaption an die Struktur der äußeren Gegebenheiten der Systemumwelt) und aktiv (die Aktivierung eines Musters entspricht dem Abruf der Repräsentation) sind. Diese kontextgebundene, dynamische Realisierung von Repräsentation, die eben nicht statisch und passiv gespeichert ist, sondern nur durch die tatsächliche Aktivierung realisiert wird, verleiht ihr einen prozeduralen Charakter. Damit verlagert sich der Ausgangspunkt für Erklärungen kognitiver Phänomene weg von den Eigenschaften der Repräsentation hin zu den Eigenschaften der Verarbeitung.

### 2.2.2.1. Die Frage der Erklärungsebene

Eine weitere Debatte betrifft die Frage der Erklärungsebene, die mit konnektionistischen Modellen verfolgt wird bzw. abgedeckt werden kann. Eine klassische Unterteilung verschiedener Erklärungsebenen für kognitive Prozesse erfolgt nach Marr (1982) für eine funktionale, eine algorithmische und eine implementationale Ebene:

- Ebene der Funktion mit funktional-abstrakter Problemanalyse (Aufgabenzerlegung in Hauptkomponenten, Bestimmung von Zweck und Ziel).
- Ebene der Algorithmen mit Beschreibung formaler Prozeduren der Aufgabenbewältigung zur Erfüllung der Funktion (Schritte der Input-Output-Transformation zur Erreichung des Ziels).
- Ebene der physikalischen Implementierung zur Bestimmung der (materiellen) Grundlage der Realisierung der Algorithmen und der Konstruktion des Systems.

Anhand dieser Ebenen werden verschiedene Zugangsweisen zur Kognition unterschieden und der Status des Konnektionismus hinsichtlich seiner Bedeutung und Aussagekraft diskutiert. Häufig werden konnektionistische Modelle als Erklärungsansätze zwischen der algorithmischen und der implementationalen Ebene positioniert. In welcher Beziehung stehen sie aber zu anderen Theorien? Inwiefern sind die verschiedenen Erklärungsansätze für Kognition nur komplementäre Ansätze, die auf bestimmte unterschiedliche Aspekte der Kognition beschränkt sind, oder aber überschneidend kompetitive Ansätze, die sich gegenüberstehen und sich möglicherweise gegenseitig ausschließen? Entgegengesetzte Positionen zu dieser Frage nehmen der eliminative und der implementationale Konnektionismus ein, eine vermittelnde Rolle vertreten hybride und integrative Ansätze (vgl.

Pospeschill, 2004; Hilario, 1997; Pinker & Prince, 1988). Die unterschiedlichen Positionen werden im folgenden kurz erläutert:

- Aus der Perspektive des implementationalen Konnektionismus besteht die eigentliche Rolle des Konnektionismus darin, ein bestehendes Symbolsystem zu implementieren: Sein einziger Wert liegt darin, eine Implementierungsmöglichkeit von Symbolsystemen zu erfassen, während er aber keinen weiteren zusätzlichen Erkenntnisgewinn für die algorithmische und die funktionale Ebene sowie für kognitive Prozesse und Strukturen leisten kann. Dieser Standpunkt wird etwa von Touretzky (1991) oder mit dem Symbolic Connectionism von Holyoak und Hummel (2000) vertreten.
- Aus der Perspektive des eliminativen Konnektionismus existiert keine symbolische Ebene der Kognition. Kognitive Prozesse und Strukturen lassen sich besser aus Prinzipien subsymbolischer Parallelverarbeitung erklären. Symbolische und konnektionistische Modelle stellen also inkompatible Erklärungsansätze mit unterschiedlichen Konsequenzen aus den angenommenen Prinzipien und Vorhersagen dar. Sie bieten substantiell unterschiedliche Erklärungen, die sich nicht aufeinander abbilden lassen. Konnektionistische Modelle sind also keine bloßen Implementierungen symbolischer Systeme, sondern leisten einen Erkenntnisgewinn, der über das hinausgeht, was symbolische Systeme leisten können. Typische Beispiele sind FFNs (Rumelhart, Hinton & Williams, 1986) oder Simple Recurrent Networks (SRNs, Elman, 1990).
- Theorien hybrider (konnektionistisch-symbolischer) Systeme vertreten die Annahme, dass beide Ansätze ihre Berechtigung besitzen: Der Konnektionismus dient der Erklärung der Funktionen peripherer Teilsysteme (v.a. Perzeption), symbolische Theorien dienen der Erklärung von zentral gesteuerten, höheren kognitiven Prozessen (etwa Denken, Problemlösen, Planen). Beide Ansätze decken also andere Aspekte kognitiver Verarbeitung ab, die für eine vollständige Erklärung kombiniert werden müssen. Vertreter hybrider Ansätze sind zum Beispiel Sun und Bookman (1994).
- Im integrativen Konnektionismus bestehen symbolische und konnektionistische Theorien gleichberechtigt als zwei Sichtweisen auf ein System kognitiver Verarbeitung. Auf Grundlage des Konnektionismus können die in symbolischen Systemen (algorithmisch) beschriebenen Gesetzmäßigkeiten und ihre Entstehung abgeleitet und begründet werden. Symbolische Theorien dienen also der Erklärung kognitiver Phänomene. Konnektionistische Theorien dienen der Erklärung der Prinzipien, nach denen symbolische Modelle operieren, und ihrer Entstehung. Diesen Standpunkt vertritt etwa Smolensky (1999) mit der Harmonic Grammar und der Optimality Theory.

Die Position des implementationalen Konnektionismus verneint also eine weitergehende Bedeutung dieses Ansatzes. Die Position hybrider Systeme reduziert die Bedeutung auf eine stark beschränkte Klasse von Aspekten. Die in dieser Arbeit im Zentrum stehende PDP-Perspektive entspricht der Position des eliminativen Konnektionismus, dessen Anspruch es ist, ein vollwertiger Erklärungsansatz zu sein – auch für höhere kognitive Prozesse.

Die Statusfrage des Konnektionismus wird die eliminative Position betreffend in einer weiteren Debatte verfolgt, die seit Ende der 1980er Jahre bis in die aktuelle Forschung hinein geführt wird: Kann mit einem subsymbolischen Modell die systematische Funk-

tionalität menschlicher Kognition erfasst werden? Auf der Suche nach dem Beweis dafür, dass konnektionistische Modelle grundsätzlich zu systematischer Verarbeitung fähig sind, beziehen sich praktisch alle Studien auf die grundlegende Kritik des Konnektionismus von Fodor und Pylyshyn (1988). Diese postulieren, dass die Auffassung des kognitiven Systems als symbolisches System unabdingbar sei, da nur dieses eine Language of Thought (Fodor, 1976) mit den Eigenschaften Produktivität, Systematizität, Kompositionalität und inferenzielle Kohärenz implementieren kann. Diese Eigenschaften erfordern kognitive Repräsentationen und Operationen, die diskrete symbolische Einheiten, eine kombinatorische Syntax und die Permanenz atomarer symbolischer Einheiten in komplexen Repräsentationen gewährleisten – analog zu Wörtern und Sätzen in einer Sprache. Fodor und Pylyshyn (1988) zufolge können konnektionistische Modelle aufgrund ihrer verteilten Repräsentationen diese Voraussetzungen nicht erfüllen. Diese könnten nur dann die erforderliche Systematizität erreichen, wenn sie ein klassisches Symbolsystem implementieren.

Diese Annahme wird aus konnektionistischer Perspektive bestritten. So erläutert van Gelder (1990), dass die Kompositionalität im Konnektionismus eigenen Regeln folgt: Entgegen der konkatenativen (d.h. diskret-syntaktisch symbolerhaltenden) Kompositionalität klassisch-symbolischer Ansätze besitzt der Konnektionismus eine funktionale Kompositionalität, die hinsichtlich ihres Potentials zumindest ebenbürtig ist. Van Gelder zeigt anhand verschiedener Beispiele, wie in konnektionistischen Modellen die systematische Kombierbarkeit von Repräsentationen und die anschließende Extrahierbarkeit der einzelnen Komponenten aus der Gesamtrepräsentation gewährleistet werden kann. Damit stellt der konnektionistische Ansatz die Bedingungen einer Language of Thought in Frage. In vielen weiteren theoretischen und empirischen Studien wurde die Kompositionalität, Produktivität und Systematizität der Repräsentation und Verarbeitung in konnektionistischen Modellen untersucht (Elman, 1991; Niklasson & van Gelder, 1994; Hadley, 1994; van Gelder & Port, 1995; van Gelder, 1998; G. F. Marcus, 1998; Bodén & Niklasson, 2000; Hadley, Rotaru-Varga, Arnold & Cardei, 2001; S. L. Frank & Haselager, 2006; S. L. Frank et al., 2009). Während sich verteilte Repräsentationen als starkes Gegenkonzept zu symbolischen Repräsentationen erweisen, herrscht hinsichtlich der Frage der Systematizität (d.h. die abstrakte Generalisierung auf neue Instanzen) bis heute Uneinigkeit betreffend ihrer genauen Leistungsfähigkeit. Dabei kristallisieren sich allerdings gewisse Grenzen und Beschränkungen des eliminativen Konnektionismus heraus.

### 2.3. Konnektionismus und Sprache

Seit der Entwicklung des Backpropagation-Algorithmus hat sich ein verstärktes Interesse am Konnektionismus als Rahmentheorie für den Gegenstand Sprache entwickelt – zur Erklärung von Erwerb, Verarbeitung und Produktion von Sprache über Störungen des Erwerbs und der Verarbeitung bis hin zur Genese und Ontogenese der Sprache. Welche Mechanismen und Strukturen befähigen uns zum Erwerb und zur Verarbeitung von Sprache? Weshalb ist Sprache genau so, wie sie ist? Und wieso zeigen wir bei der Verwendung von Sprache die spezifischen beobachtbaren Eigenschaften? Konnektionistische Netzwerke dienen als Modell zur Simulation von Sprachverarbeitungsaspekten. Darüber hinaus ermöglichen sie die Überprüfung von Annahmen über die Eigenschaften unserer Sprachfä-

higkeit und zur Formulierung alternativer Annahmen. Die konnektionistische Perspektive hat in den vergangenen rund 25 Jahren zu umfangreichen Neuinterpretationen und zur kritischen Hinterfragung klassischer Annahmen über die Sprachverarbeitung geführt. Im Zusammenhang mit Modellierungen konkreter sprachlicher Verhaltensdaten wird auch von einer konnektionistischen Psycholinguistik (Christiansen & Chater, 2001b) gesprochen. Wir heben in den nächsten Abschnitten einige Debatten hervor, die im Verlauf der Arbeit wieder auftauchen werden.

**Spracherwerb** Eine grundlegende Debatte, in der die Gegensätzlichkeit von Konnektionismus und Generativer Grammatik – speziell hinsichtlich der Annahme einer angeborenen Universalgrammatik – kulminiert, betrifft die Frage des Erwerbs und der Angeborenheit sprachlicher Strukturen und Fähigkeiten: die Nature-Nurture-Debatte (für eine Übersicht vgl. E. Bates et al., 1999). Vor allem begründet über das Argument der Poverty of the Stimulus (Chomsky, 1975) wird angenommen, dass Sprache nicht induktiv gelernt werden kann, sondern deduktiv auf der Grundlage angeborenen sprachspezifischen Wissens erworben werden muss (vgl. Fanselow & Felix, 1987). Spracherwerb erfolgt Theorien dieser Tradition folgend als Deduktion und Parameteranpassung (z.B. Pinker, 1994; Lightfoot, 1999). Der Konnektionismus stellt dieser Annahme eine empiristische Position entgegen, nach der kein angeborenes sprachspezifisches Wissen gegeben sein muss, und Sprache rein induktiv anhand assoziativer Mechanismen aufgrund statistischer Eigenschaften des Inputs erworben werden kann (z.B. Elman et al., 1996; J. Lewis & Elman, 2001a, 2001b).

**Aufhebung klassischer Dichotomien** Hinsichtlich des Erwerbs und der Verarbeitung von Sprache werden des Weiteren viele fundamentale Dichotomien in Frage gestellt, die in vielen linguistischen und psycholinguistischen Theorien sowie symbol- und produktionsystemorientierten kognitiven Modellen mitunter als axiomatisch für die menschliche Sprachverarbeitung angenommen wurden. Dies betrifft beispielsweise die klassische Trennung von Kompetenz und Performanz, wie sie unter anderem in der Generativen Grammatik postuliert wird (Chomsky, 1965). Dieser Dichotomie zufolge verfügt der Mensch über ein absolutes und unbeschränktes grammatikalisches System (die Kompetenz), mit dem er alle beliebig komplexen, in seiner Sprache wohlgeformten Äußerungen verstehen, produzieren und in ihrer Grammatikalität beurteilen kann. Dem steht die Performanz gegenüber, die das sprachliche Verhalten, das auf der Grundlage der Kompetenz durch das beschränkte kognitive und motorische System hervorgebracht wird, beschreibt (vgl. Fanselow & Felix, 1987). Im Konnektionismus wird diese Dichotomie aufgehoben: Es gibt nur noch sprachliches Verhalten, und die sprachliche Kompetenz ist direkt durch die kognitiven Eigenschaften des Systems bedingt (z.B. Christiansen & Chater, 1999b). So weisen die Modelle eine starke Nähe zu gebrauchsbasierten (usage-based grammar, vgl. Bybee & McClelland, 2005) und statistischen (z.B. Wells, Christiansen, Race, Acheson & MacDonald, 2009) Ansätzen der Sprache auf, die einen starken Einfluss der Sprachverwendung und der distributionalen Eigenschaften des sprachlichen Inputs auf die kognitive Repräsentation und Verarbeitung von Sprache annehmen.

Des Weiteren wird damit zusammenhängend die Unterscheidung von grammatischem (prozeduralem) Wissen in Form von Regeln und von statisch gespeichertem (deklarativem) Symbolwissen, etwa in Form eines mentalen Lexikons, abgelehnt. Diese Trennung lässt

sich mit der Auflösung der Grenzen zwischen Verarbeitung und Repräsentation nicht mehr aufrechterhalten. An die Stelle von Regeln und Ausnahmen treten komplexe Assoziationen innerhalb eines Systems, in dem Aktivierungen durch multiple Constraints bestimmt werden (z.B. McClelland & Elman, 1986; McClelland et al., 1989). Lexikalisches Wissen wird nicht mehr passiv und statisch gespeichert und abgerufen. Vielmehr fungieren lexikalische Einheiten als Operatoren bei der kontextsensitiven Aktivierung spezifischer Netzwerkzustände (vgl. Elman, 2004, 2009).

**Sprachverarbeitung** Durch die massive Parallelverarbeitung implementieren konnektionistische Netzwerke Prinzipien des Multiple-Constraint-Ansatzes (vgl. MacDonald & Seidenberg, 2006). Bei der Verarbeitung des Inputs werden mehrere Informationsquellen integriert und wirken parallel als Beschränkung der möglichen Zustände des Systems. Damit stehen konnektionistische Modelle als interaktive Ansätze den modularen Ansätzen gegenüber, die eine gekapselte Verarbeitung verschiedener Aspekte und mitunter eine serielle Schaltung der Verarbeitung annehmen. Auf Grundlage der parallelen Verarbeitung und der verteilten Repräsentation erfolgt die Aufhebung der angenommenen Trennung von generellem und spezifischem Wissen (McClelland & Rumelhart, 1985) und der Type-Token-Distinktion (Elman, 1990). Auch die Unterscheidung zwischen regelmäßigen Fällen und Ausnahmen (vgl. dazu etwa die Past-Tense-Debatte, Rumelhart & McClelland, 1986a) wird aufgegeben – einschließlich der Annahme der modularen Verarbeitung der unterschiedlichen Formen des Inputs. Konnektionistische Modelle stellen dieser Annahme einen Ansatz gegenüber, in dem regelmäßige und unregelmäßige Formen innerhalb eines einzigen Systems erfasst werden können (für aktuellere Zusammenfassungen der Past-Tense-Debatte vgl. McClelland & Patterson, 2002; Pinker & Ullman, 2002).

Bei der Suche nach Erklärungen für empirisch beobachtete Phänomene der Sprachverarbeitung und nach Vorhersagen des sprachlichen Verhaltens werden konnektionistische Modelle im Sinne einer konnektionistischen Psycholinguistik (Christiansen & Chater, 2001b) herangezogen, um in den verschiedensten linguistischen Domänen – Phonetik, Lexik, Syntax, Semantik, Sprachproduktion und -rezeption, Lesen – Leistungen, Eigenschaften und Fehler zu simulieren. Im Zentrum steht die Deckung zu Daten und Performanzverhältnissen, wie sie in psycholinguistischen Studien erhoben werden, unter anderem Lese- und Reaktionszeiten, Interpretation und Verständnisleistung, Fehlleistungen, sowie Blick- bzw. Aufmerksamkeitslenkung (einen Überblick zu entsprechenden konnektionistischen Modellen bieten Christiansen & Chater, 1999a, 2001b, 2001a). Es geht dabei nicht mehr nur um die Frage, ob konnektionistische Modelle ohne die Realisierung klassischer Axiome der Sprachverarbeitung und ohne die Annahme angeborenen Wissens komplexe sprachliche Strukturen erwerben und verarbeiten können. Es geht auch darum, im Detail empirisch-psychologische Daten zu modellieren. Dabei stehen andere Fragen und Kriterien der Evaluation der Modelle im Vordergrund, unter anderem: (a) die Frage der Skalierbarkeit, d.h. die Modellierbarkeit realistischer Datenmengen und -strukturen; (b) die Frage der Architektur, etwa, welche Module für Teilaspekte angenommen werden sollen; (c) die Frage des vorgegebenen Wissens, zum Beispiel, wieviel strukturelles Wissen durch die Gestaltung des Inputs vorgegeben werden darf; und (d) die Frage der Vergleichbarkeit mit empirischen Daten und der Operationalisierung konnektionistischer Performanzdaten (vgl. Christiansen & Chater, 2001b).

**Sprachentwicklung** Um den Kreis zu schließen, greifen wir wieder den Aspekt der Annahme angeborenen sprachspezifischen Wissens in einer Universalgrammatik auf. Im Rahmen konnektionistischer Modelle wurden verschiedene Erklärungsansätze zur Genese und Entwicklung von Sprache, zur Entstehung der menschlichen Befähigung für Sprache sowie zu Eigenschaften sprachlicher Typologie entwickelt. Aufgrund konnektionistischer Eigenschaften und Prinzipien werden Erklärungen dafür gesucht, warum Sprache so ist, wie sie ist. Damit werden Ansätze in Frage gestellt, die auf angeborenem Wissen, vorgegebenen Strukturen (im Sinne genetisch eingeschriebener sprachübergreifender Prinzipien und Parameter) und deduktivem Lernen basieren. In diesem Zusammenhang wurde Sprache als dynamisches System (Elman et al., 1996; Elman, 2004; van Gelder & Port, 1995) bzw. als emergente Struktur (Elman, 1998; MacWhinney, 1998) aus der Interaktion von System und Umwelt beschrieben. Eine evolutionstheoretische Perspektive entwickeln Christiansen und Chater (2008) in ihrem Artikel „Language as shaped by the brain“. Diesem Ansatz zufolge entwickelte sich im Laufe der Evolution keine spezialisierte kognitive Fähigkeit oder neuronale Komponente, die Sprache hervorbringen konnte, vielmehr bestimmt die Struktur des Gehirns und seine kognitiven Eigenschaften, welche Struktur unsere Sprache annimmt. Sprache wird also im Sinne eines Systems verstanden, das sich ähnlich einem Organismus dem Selektionsdruck der Umweltaforderungen sowie unserer kognitiven Eigenschaften anpasst (zu weiteren Aspekten der Evolution von Sprache vgl. Christiansen et al., 2002; Christiansen & Kirby, 2003). In verschiedenen Simulationsstudien wurde versucht, Frequenzen und Beschränkungen typologischer Merkmale zu modellieren. Die Beschränkungen möglicher Sprachen wurden dabei anhand allgemeiner sprachspezifischer Prinzipien erklärt, vor allem durch die Lernbarkeit verschiedener sprachlicher Strukturen für assoziativ-induktive Lernmodelle (u.a. Christiansen & Devlin, 1997; S. Levy, Melnik & Pollack, 2000; Christiansen et al., 2002; van Everbroeck, 2003).



## 3. Sprachverarbeitung in Simple Recurrent Networks

Wie bereits im vorhergehenden Kapitel angedeutet, wurden Simple Recurrent Networks (SRNs, Elman, 1990) vielfach erfolgreich eingesetzt, um menschliche Satzverarbeitung – und auch allgemeiner Sequenzverarbeitung – zu modellieren. Mittlerweile sind SRNs innerhalb des konnektionistischen Paradigmas die klassische Architektur, um temporale bzw. sequenzielle Eigenschaften zu erfassen. Sie werden in verschiedenen komplexeren Architekturen eingesetzt, um diesen Aspekt der Sprachverarbeitung zu integrieren. Dabei legt das Modell Prinzipien und Mechanismen der Verarbeitung, Repräsentation und des Erwerbs nahe, die von einer breiten Klasse anderer Modelle und vor allem von konventionellen repräsentations- und regelbasierten Modellen abweichen.

In diesem Kapitel werden wir die SRN-Architektur besprechen. Zunächst werden wir die Grundfunktionen der Sequenzverarbeitung in SRNs sowie allgemeinen Verarbeitungseigenschaften wie etwa die Klassifikation sprachlicher Strukturen (Derivation syntaktischer Klassen und semantischer Distinktionen, vgl. Elman, 1990, 1991, 2004) erläutern. Dabei werden wir verschiedene Debatten wie die Erwerbsfähigkeit für verschiedene sprachliche Strukturen und herausgehoben die Frage der Systematizität beleuchten. Anschließend werden wir einige Studien näher beleuchten, in denen spezifische Eigenschaften dieser Architektur bei der Verarbeitung von Wortsequenzen und der Aufgabe der Wortvorhersage identifiziert werden. Besprochen werden wir unter anderem Effekte der Komplexität (z.B. Christiansen & Chater, 1999b), der Frequenz und der Regularität (MacDonald & Christiansen, 2002; Wells et al., 2009) und der Interferenz des lokalen Kontexts (Konieczny & Döring, 2003; Konieczny, Ruh & Müller, 2009; Weldle, Baumann et al., 2009). In erster Linie geht es darum, Erklärungsmuster zu identifizieren, die für die Eigenschaften von SRNs bei der Sequenzverarbeitung herangezogen werden. Diese gilt es auf menschliche Sprachverarbeitung als deskriptives, erklärendes und prädiktives Modell anzuwenden. Anders als explizite symbolische Modelle sind konnektionistische Modelle selbst als Gegenstand der Untersuchung zu verstehen, deren Funktionsweise erst nachvollzogen werden muss. Deshalb werden wir in Exkursen die Evaluation der Netzwerk-Performanz und die hier gängigen Fehlermaße sowie relevante multivariate statistische Verfahren zur Analyse der internen Strukturen besprechen.

### 3.1. Sequenzielle Verarbeitung: Modellierung mit SRNs

SRNs sind eine Erweiterung einfacher Feedforward-Netzwerke (FFNs). Sie verfügen über einen Input-, einen Output- und einen Hidden-Layer sowie über eine beliebige Anzahl weiterer versteckter Kompressions-Layer. Die Besonderheit der Architektur gegenüber

einfachen FFNs ist die Rekurrenz des Hidden-Layers: Dessen Aktivierung wird in einem zusätzlichen Kontext-Layer zwischengespeichert, das die Aktivierung zum jeweils nächsten Zeitschritt anreichert. Diese Speicher-Komponente stattet das Netzwerk mit einem dynamischen Gedächtnis für vergangene Ereignisse der Sequenz – den sequentiellen Kontext des aktuellen Ereignisses – aus.

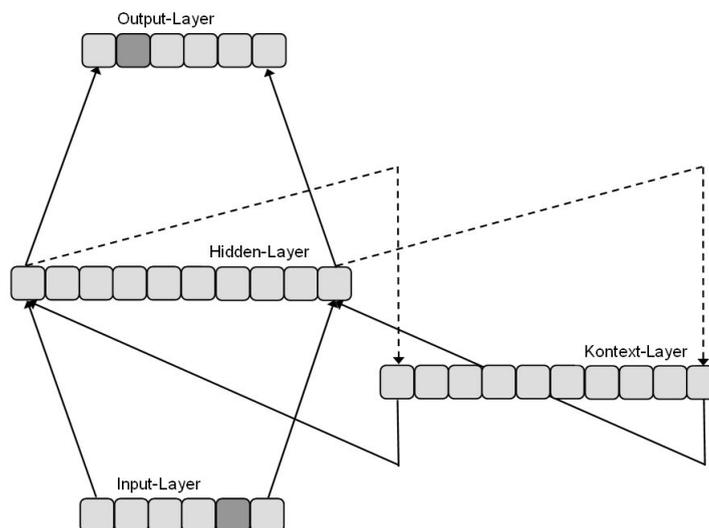


Abbildung 3.1.: Architektur eines SRNs. Aktivierung fließt vom Input- und Kontext-Layer zum Hidden-Layer und weiter zum Output-Layer. Die Aktivierung des Hidden-Layers wird über fixierte Gewichte auf den Kontext-Layer kopiert. Der Fehler wird anhand des Algorithmus Backpropagation zugewiesen. Der Kontext-Layer dient als dynamisches Gedächtnis des Netzwerks. Es speist zu jedem Zeitschritt die Aktivierung des Hidden-Layers aus dem vorhergehenden Zeitschritt in die neue Aktivierung ein und setzt damit die Aktivierung des Hidden-Layers in seinen sequenziellen Kontext.

Wir verfolgen den Aktivierungsfluss in der folgenden Abbildung des SRN-Modells (Abb. 3.1): Zu jedem Zeitschritt fließt Aktivierung über vollständige Verknüpfungen nicht nur vom Input-Layer, sondern auch vom Kontext-Layer zum Hidden-Layer und von dort aus weiter zum Output-Layer. Auf umgekehrtem Weg erfolgt die Fehlerzuweisung und Gewichts Anpassung wie in FFNs anhand des Backpropagation-Algorithmus. Bei der vorwärtsgerichteten Propagierung der Aktivierung wird zusätzlich das Aktivierungsmuster des Hidden-Layers über eine Kopierverknüpfung auf den Kontext-Layer kopiert: Jede Unit des Hidden-Layers ist mit genau einer Unit des Kontext-Layers über ein fixes Verbindungsgewicht der Stärke 1 verknüpft. Zudem verfügen die Units des Kontext-Layers über eine lineare Aktivierungsfunktion – dadurch ist sichergestellt, dass ein identisches Aktivierungsmuster übertragen und gespeichert wird. Durch diesen Mechanismus wird zu jedem Zeitschritt der Inhalt des Hidden-Layers aus dem vorhergegangenen Zeitschritt in die aktuelle Aktivierung des Hidden-Layers eingespeist und dieser aktuelle Inhalt des Hidden-Layers zwischengespeichert. Dadurch, dass der jeweils neu zu speichernde Inhalt bereits mit Kontextinformationen angereichert ist, werden latente Informationen nicht nur aus einem, sondern aus mehreren Zeitschritten verfügbar gemacht. Das Kontext-Layer funktioniert damit als dynamisches Gedächtnis des Netzwerks, über das zu jedem Zeit-

schritt zusätzliche Informationen über den sequenziellen Kontext der aktuellen Aktivierung in die Transformation zur Target-Aktivierung eingehen.<sup>1</sup>

Wir führen an dieser Stelle die Unterscheidung zweier grundlegend unterschiedlicher Verwendungsweisen von SRNs ein: (a) SRNs mit Vorhersage-Aufgabe (vgl. Kap. 3.1.2), bei der das jeweils folgende Item einer Sequenz vorhergesagt wird, mit dem Ziel, jedem Input kontextabhängige Übergangswahrscheinlichkeiten zum Folge-Item zuzuweisen, und (b) SRNs mit Assoziations- oder Komplettierungs-Aufgabe, bei dem eine Sequenz aufeinanderfolgender Items auf eine bereits vollständige statische Struktur (oder ein statisches Segment einer umfangreicheren Struktur) gemappt werden, mit dem Ziel, jedem Input eine kontextabhängige Rolle im Constraint-Satisfaction-Prozess bei der Konstruktion der vollständigen Struktur zuzuweisen (vgl. Kap. 4). Bei erstgenannter Aufgabe wird keine vollständige persistente Struktur der Sequenz aufgebaut, vielmehr wird für jede Position im Satz die Entropie für die Projizierung zukünftiger Positionen bzw. die Verteilung der Wahrscheinlichkeiten verschiedener Wortklassen im syntaktischen Kontext der Position bestimmt. Diese Flüchtigkeit der Bestimmung für jeweils nur eine Position ohne Aufbau einer syntaktischen Struktur mit Bindung der Elemente der Sequenz entspricht nicht der klassischen Vorstellung eines Parsingmodells und kann nicht als solches verstanden werden. Wesentlich näher an klassischen Parsingmodellen ist die Umsetzung der zweitgenannten Aufgabe, bei der in Abhängigkeit von der eingehenden Sequenz eine persistente Struktur des Satzes oder eine Repräsentation des gesamten Satzinhalts aufgebaut werden soll.

#### 3.1.1. Alternative rekurrente Architekturen

Neben SRNs gibt es eine ganze Reihe weiterer rekurrenter konnektionistischer Modelle, unter anderem rekursive auto-assoziative Netzwerke sowie Architekturen mit sogenanntem Reservoir Computing. Da sie im späteren Verlauf eine geringe Rolle spielen, werden wir sie an dieser Stelle nur in aller Kürze erwähnen.

Die Architektur des Recursive Auto-Associative Memory (RAAM, Pollack, 1990) ermöglicht die Komprimierung von mehreren Input-Items sowie die Dekomposition solcher komprimierter Repräsentationen. In der ursprünglichen Form sind RAAM-Netzwerke als dreischichtige FFNs realisiert, die auto-assoziativ einen Input auf einen identischen Output abbilden. Der Input-Layer unterteilt sich in zwei Segmente, deren Informationen im Hidden-Layer zu einem Zeitschritt in eine gemeinsame Repräsentation überführt werden. Diese Repräsentation wird rekursiv in den Input-Layer eingespeißt und stellt dort die Informationen eines der beiden Segmente für den nächsten Zeitschritt. Zu jeder neuen Information im einen Segment wird also über das zweite Segment die Kontext-Information der bisherigen Gesamtinformation (die komprimierte Repräsentation aus dem Hidden-Layer) aus den vorhergehenden Zeitschritten zugefügt. Das Resultat ist eine komplexe Repräsentation, die die gesamte sequenzielle Information beinhaltet.<sup>2</sup> Über diesen Mecha-

---

<sup>1</sup>Einen ähnlichen Ansatz verfolgt das etwas frühere Modell von Jordan (1986), in dem die Rekursion vom Output in das Hidden-Layer erfolgt. Die Aktivierung des Hidden-Layers erfolgt also im Kontext der vorhergegangenen Output-Performanz.

<sup>2</sup>Inzwischen wurden mehrere Erweiterungen und Verbesserungen dieser Architektur vorgeschlagen. Adamson und Damper (1999) schlagen mit sogenannten B-RAAMs (Bi-coded RAAMs) eine Erweiterung vor, die erweiterte Kontext-Informationen nutzt und damit bessere und stabilere Verarbeitungseigenschaften

nismus können in dieser Architektur Sequenzen mit hierarchischer Struktur und variabler Größe verarbeitet und in einer komprimierten Struktur holistisch repräsentiert werden. Auf dieser Grundlage entwickelt zum Beispiel Berg (1992) mit dem XERIC-Parser ein konnektionistisches Modell syntaktischer Verarbeitung. In diesem werden mit RAAMs auf X-Bar-Strukturen basierende komprimierte Satzrepräsentationen erzeugt.

Eine weitere Klasse bilden Netzwerke mit sogenanntem Reservoir Computing. Zwei Hauptvertreter davon sind Echo State Networks (ESNs, H. Jaeger, 2003) und Liquid State Machines (LSMs, Maass, Natschläger & Markram, 2002). Diese Netzwerke verwenden eine rekurrente Komponente mit fixierten Gewichten und sparsamer willkürlicher Verknüpfung von Units. Sie bauen im Unterschied zu SRNs auf dem Prinzip auf, dass diese Komponente nicht trainiert werden muss, sondern ein unspezifisches dynamisches Gedächtnis bildet, auf den der Input in einer höherdimensionalen Repräsentation abgebildet wird. Trainiert werden muss nur ein Dekodier-Netzwerk, das die internen Zustände des Netzwerks interpretiert. S. L. Frank (2006b, S. 287) beschreibt die Eigenschaften des Netzwerks bezüglich des rekurrenten Layers als „combining a large short-term memory with a small long-term memory capacity“. In einer vergleichenden Simulationsstudie untersucht er die Kapazität von SRNs und ESNs hinsichtlich ihrer Systematizität als Satzverarbeitungsmodell. S. L. Frank (2006b) stellt dabei als Vorteil der Gedächtniseigenschaften von ESNs fest, dass sie es ermöglichen, ein sehr großes rekurrentes Layer zu verwenden, ohne die Gefahr des bei SRNs beobachteten Overfittings einzugehen. Ein dem Reservoir Computing ähnliches Prinzip verfolgen auch Dominey und Ramus (2000) mit ihren Temporal Recurrent Networks (TRNs). Diese Variante verfügt über zusätzliche direkte nicht-rekurrente Verknüpfungen und über fixierte rekurrente Gewichte. Das heißt, im Unterschied zu SRNs sind die Gewichte vom und zum Kontext-Layer in beide Richtungen fixiert (also nicht trainierbar).

All diese alternativen Architekturen zur Verarbeitung von Sequenzen haben ihre eigenen Eigenschaften, Stärken und Schwächen, deren Besprechung und Anwendung jenseits des Rahmens dieser Arbeit liegt. Die klassische Architektur, die am häufigsten und vielseitigsten für sprachliche Fragestellungen umgesetzt wurde, bleibt das SRN. Dessen Funktionsweise werden wir in den folgenden Abschnitten näher betrachten.

### 3.1.2. Sequenzfassung durch Prädiktion

In der Studie „Finding Structure in Time“ stellt Elman (1990) verschiedene Beispiele vor, die die Vorhersage-Aufgabe und seine Eigenschaften illustrieren. Die Modelle lernen, das jeweils nächste Element einer Sequenz und seine assoziierten Eigenschaften auf der Grundlage der bisher verarbeiteten Sequenz vorherzusagen. Dieser Prozess wird auch als Projektion bezeichnet und wurde für verschiedene Aspekte der Sprachverarbeitung belegt. Die Modelle nutzen für diese Aufgabe den Kontext der bisher verarbeiteten Sequenz, der durch die rekurrente Schicht zur Verfügung gestellt wird.

---

ten für komplexe sequenzielle Strukturen aufweist. S. Levy et al. (2000) stellen die Architektur Infinite RAAM (IRAAM) vor, eine Fusion von rekurrenten Netzwerken und Fraktalgeometrie, die potentiell dazu in der Lage sein soll, infinite Rekursion zu erfassen.

Die Erfassung eines Gegenstands als inhärent sequenzielles Probleme bietet eine neue Perspektive auf den Gegenstand. Diese Änderung der Eigenschaften zeigt Elman anhand einer Variante des XOR-Problems, bei der Bedingungs- und Ergebniszustände als binäre Input-Sequenz gegeben sind. Die sequenzielle Verarbeitung über die Zeit mit der Aufgabe der Vorhersage des Folge-Items erfordert einen neuen Lösungsmechanismus, bei dem sich die Bedingungssteile als Kontext für die Lösung dienen. Da jeweils nur für die abhängige Ergebniszustände eine sinnvolle Vorhersage getroffen werden kann (die Bedingungsstände sind unabhängig), muss das SRN auch Wissen über die temporale Struktur des Inputs erwerben. Ein weiteres Beispiel untersucht die Verarbeitung von Wörtern als kontinuierliche Sequenz von Buchstaben. Aus der Perspektive mit sequenzieller Darbietung und Vorhersage-Aufgabe entwickelt sich ein Erklärungsansatz für das Segmentierungsproblem im Erstspracherwerb – das Problem des Erkennens von Wortgrenzen in einem kontinuierlichen Lautstrom – im Sinne von Unterschieden der Entropie: Sequenzgrenzen bzw. Wortgrenzen werden durch einen plötzlichen Anstieg des Vorhersagefehlers für den jeweils folgenden Buchstaben markiert.

Diesen Ansatz erweitert Elman zu einem Modell der Verarbeitung einfacher Sätze als Vorhersage des jeweils nächsten Wortes. Dem Netzwerk werden dabei Sätze als Wortsequenzen Wort für Wort dargeboten. Als Target liegt am Output das jeweils nächste Wort der aktuellen Sequenz an – d.h. das Wort, das zum nächsten Zeitschritt als Input dargeboten wird. Um diese Aufgabe sinnvoll zu bewältigen, muss das Netzwerk über die Kookkurrenzen von Wörtern über verschiedene Sätze hinweg die relevanten Zusammenhänge zwischen den Wörtern aus dem Datenstrom extrahieren und enkodieren. Was das Netzwerk dabei lernt, ist nicht die Vorhersage eines einzelnen Folgeworts, sondern die Vorhersage aller Wörter, die für die jeweilige Position möglich und grammatikalisch korrekt sind und damit die Kookkurrenz-Constraints erfüllen. Das folgende Beispiel demonstriert die Aufgabe der Wortvorhersage und die mögliche Performanz nach dem Training. Für Satz 1 werden alle Wörter des Satzes Wort für Wort in aufeinanderfolgenden Schritten dargeboten. Im ersten Schritt erhält das Netzwerk das erste Wort *Der* als Input. Als Target wird das nächste Wort der Sequenz angelegt, hier das Wort *Junge*. Im nächsten Schritt erhält das Netzwerk das folgende Wort *Junge* als Input, Target ist das wiederum folgende Wort *sieht*. So erhält das Netzwerk Schritt für Schritt alle Wörter des Satzes.

- (1) Der Junge sieht das Mädchen EOS ...
  - a. Input: Der → Target: Junge  
Prediction: Junge/Mann/Polizist/kleine/böse/etc.
  - b. Input: Junge → Target: sieht  
Prediction: sieht/kratzt/schläft/will/der/und/etc.
  - c. Input: sieht → Target: das  
Prediction: das/den/die/einen/eine/immer/gerade/etc.
  - d. Input: das → Target: Mädchen  
Prediction: Mädchen/Haus/schöne/bunte/etc.
  - e. Input: Mädchen → Target: EOS  
Prediction: EOS/neben/mit/und/obwohl/etc.
  - f. Input: EOS → Target: ...  
Prediction: Der/Die/Als/Weil/etc.

Nachdem das Netzwerk mit einem Trainingsset trainiert wurde, das aus vielen verschiedenen Sätzen besteht, hat es gelernt, dass nach einem Wort eine bestimmte Klasse anderer Wörter folgen kann. In einem Test-Set kann das Netzwerk auf jeder Position innerhalb einer Sequenz die möglichen Folgewörter vorhersagen. Nach dem Artikel *Der* können zum Beispiel Nomina mit den grammatikalischen Eigenschaften Kasus Nominativ, Numerus Singular und Genus Maskulinum folgen; oder aber Adjektive mit Übereinstimmung von Kasus und Numerus. Auf der Position des Verbs *sieht* sollte es berücksichtigen, dass es bereits ein Objekt im Nominativ gesehen hat, und der Valenz des Verbs zufolge nur Objekte im Akkusativ vorhergesagt werden dürfen; bei der im Deutschen ebenfalls korrekten Wortstellung OVS müssten bei der Vorhersage die umgekehrten Einschränkungen berücksichtigt werden. Außer bestimmten und unbestimmten Artikeln für ein Nomen im Akkusativ könnten an entsprechender Stelle zum Beispiel auch Adverbien folgen. Welche Wörter tatsächlich vorhergesagt werden und welche Wahrscheinlichkeit sie dabei erhalten, hängt vor allem vom Korpus und dessen lexikalischen, grammatischen und distributionalen Eigenschaften ab.

Wir werden in den nachfolgenden Unterkapiteln mehrere erfolgreiche Simulationen und deren Eigenschaften näher betrachten, die diese Architektur mit der beschriebenen Trainingsmethode verwendet haben. Vorab können wir aber bereits mehrere sinnvolle und (neuro-)psychologisch plausible Eigenschaften des Modells feststellen, die aus der Implementierung der Vorhersage-Aufgabe resultieren:

- Die Aufgabe der Wortvorhersage zwingt das Netzwerk dazu, geeignete interne Repräsentationen aufzubauen, die Eigenschaften der sequenziellen Struktur und Abhängigkeiten zwischen den Elementen der Sequenzen auf sehr abstraktem Niveau abbilden.
- Die Fähigkeit zur Vorhersage in Abhängigkeit der vorausgegangenen Wörter der Sequenz legt nahe, dass das System in irgendeiner Form grammatisches Wissen (im Sinne von Regelmäßigkeiten, die der Sprache zugrunde liegen) extrahiert hat und anwenden kann.
- Die Vorhersage-Aufgabe liefert möglicherweise auch eine Erklärung für grundlegende Fragen des Erstspracherwerbs, so etwa als Ansatz für das Problem der Lautstromsegmentierung (Identifikation von Wortgrenzen) anhand der Entropie der möglichen Folgeklänge in einer Sequenz, oder als Ansatz für den Erwerb syntaktischer Strukturen und Abhängigkeiten (Parsing als Effekt eines Vorhersage-Mechanismus).
- Die Aufgabe kann als selbstüberwachtes Lernen bezeichnet werden (vgl. Elman, 1991, S. 201). Dies ist wünschenswert, da das Netzwerk damit nur Informationen bekommt, die in der Umwelt gegeben sind. Es bekommt also keine expliziten Informationen über Strukturen, Abhängigkeiten und Kategorien vorgegeben und ist damit relativ theorieneutral (Elman, 1995). Dabei rückt der Ansatz ab von Modellen mit überwachtem Lernen, deren vorgegebene Zielstrukturen erklärt und gerechtfertigt werden müssen. Zugleich rückt er in die Nähe von auto-assoziativen Modellen sowie hinsichtlich der selbstorganisierenden Entwicklung interner Strukturen hin zu unüberwachten Modellen (vgl. Kap. 2.2.1.1).
- Die Vorhersage bzw. Antizipation von Eigenschaften kommender Wörter ist ein kognitionspsychologisch plausibler Mechanismus, für den es in verschiedenen psy-

chologischen Domänen Befunde gibt. Hinweise auf entsprechende Mechanismen in der Sprachverarbeitung stammen aus behavioralen (z.B. Altmann, 1999; Altmann & Kamide, 1999; Konieczny, 2000; Konieczny & Döring, 2003; T. F. Jaeger, Fedorenko & Gibson, submitted) und neurofunktionalen Studien (DeLong, Urbach & Kutas, 2005; Otten & van Berkum, 2008).

- Die Überlagerung der Informationen in den Aktivierungen der internen Layer über mehrere Zeitschritte hinweg stellt einen inhärenten Mechanismus für Gedächtniseffekte wie Aktivierungsabnahme, interferierende Überlagerung oder Reaktivierung von Informationen dar.

#### 3.1.3. Eigenschaften der Sprachverarbeitung in SRNs

Was erwerben SRNs und welche internen Strukturmerkmale entwickeln sich, um die sinnvolle Verarbeitung des Inputs gewährleisten zu können? Welche sprachlichen Regelmäßigkeiten können sie überhaupt erwerben und verarbeiten, und welche statistischen Regelmäßigkeiten nutzen sie dabei? Wie ist es um die Generalisierungsfähigkeit und die semantische Systematizität bestellt? In den folgenden Abschnitten diskutieren wir die Leistungsfähigkeit und die Grenzen von SRNs als Modell menschlicher Sprachkompetenz.<sup>3</sup> Prädiktive Aspekte der Netzwerkperformanz für empirische Fragestellungen der psycholinguistischen Forschung besprechen wir im darauffolgenden Unterkapitel 3.2.

Hinsichtlich der Frage der Lern- und Leistungsfähigkeit stehen vor allem die frühen SRN-Studien immer wieder in Beziehung zur Nativismus-Empirismus-Debatte: Ist sprachliches Wissen – Grammatik und Strukturbildungsprinzipien – angeboren, wie von Vertretern der Generativen Grammatik mit Language Acquisition Device und Universalgrammatik postuliert wird (vgl. z.B. Chomsky, 1980, 1981; Fanselow & Felix, 1987; Pinker, 1994; Piattelli-Palmarini, 1994)? Wenn ja, was genau ist bzw. welche Aspekte der Sprache sind in welchem Ausmaß angeboren und welche Rolle spielt die Umwelt und der sprachliche Input für den erfolgreichen Erstspracherwerb? Oder ist überhaupt kein sprachliches Wissen angeboren – beruhen unsere sprachlichen Fähigkeiten auf allgemeinen kognitiven Fähigkeiten und ihrer Interaktion mit der sozialen und der sprachlichen Umwelt?

Die konnektionistischen Ansätze in der PDP-Tradition bilden eine Gegenposition zu den starken nativistischen Annahmen, denn der Erwerb syntaktischer Strukturen und Regelmäßigkeiten beruht ausschließlich auf der Interaktion eines allgemeinen Systems mit spezifischen sprachlichen Stimuli und entwickelt dabei selbstorganisiert die für die Verarbeitung erforderlichen internen Strukturen.<sup>4</sup> Dies bedeutet aber nicht, dass im Konnek-

---

<sup>3</sup>Das Thema ist hierbei die Erwerbsfähigkeit: Was können SRNs überhaupt erwerben, und für welche Strukturen weist die Performanz auf Probleme der Architektur hin? Der Vergleich der Eigenschaften von SRNs im Verlauf des Erwerbs mit Daten aus der empirischen Spracherwerbsforschung bleibt dabei zweitrangig. Ein Überblick über Erklärungen des Konnektionismus zu spezifischen Entwicklungsmustern des kindlichen Spracherwerbs und aktuelle Desiderate der konnektionistischen Forschung in diesem Bereich bespricht Elman (2005).

<sup>4</sup>In verschiedenen Abstufungen finden sich in vielen anderen linguistischen Forschungsrichtungen weitere entgegengesetzte Positionen zu starken nativistischen Annahmen, so etwa in verschiedenen Ausprägungen der Konstruktionsgrammatik (vgl. Tomasello, 2005, 2006) und in verschiedenen Theorien des Spracherwerbs (etwa Behaviorismus, Kognitivismus, Interaktionismus).

tionismus die Annahme jeglicher angeborener Sprachfähigkeit abgelehnt wird. Vielmehr wird in Frage gestellt, welcher Natur die angeborenen Strukturen sind (z.B. Verlagerung auf Eigenschaften der neuronalen Funktionalität, Elman et al., 1996; Elman, 1999b). Weitere theoretische Diskussionen beinhalten auch die Frage, weshalb nur Menschen eine Grammatik und eine voll ausgebildete Sprache entwickelt haben und erwerben können (z.B. Ellis, 1998; Elman, 1999b; Christiansen et al., 2002; Elman, 2005; Christiansen & Chater, 2008).

Viele vor allem frühe Studien mit SRNs sind in dieser Hinsicht als Machbarkeitsstudien aufgebaut, mit denen untersucht werden sollte, welche Strukturen von SRNs erworben werden können. Von besonderem Interesse waren dabei Aspekte natürlicher Sprache, von denen behauptet wurde, dass sie nicht durch rein assoziative induktive Mechanismen und endliche Automaten erfasst werden können – oder für die aufgrund logischer Argumente wie etwa dem Poverty-of-the-Stimulus-Argument mit dem zentralen Problem fehlender negativer Evidenzen (vgl. Chomsky, 1980) die Notwendigkeit angeborener Strukturen angenommen wurden. Das Fehlen negativer Evidenzen ist insofern ein Problem, als das Kind beim Spracherwerb nur positive Beispiele oder aber inkorrekte oder unvollständige Sätze hört, die unkorrigiert geäußert und akzeptiert werden. Niemals hört es aber explizite Beispiele für falsche Sätze, die in der korrekten Grammatik nicht verwendet werden dürfen und als falsche Strukturen kommentiert werden. Zudem wurde behauptet, dass die von den Kindern selbst geäußerten falschen Sätze selten oder sogar überhaupt nicht korrigiert werden. Für ausschließlich positive Beispiele wird angenommen, dass sie nicht für einen rein induktiven Erwerb von Sprache genügen. Um den erfolgreichen Erwerb aus ausschließlich positiven Beispielen zu gewährleisten, muss nach dieser Argumentation angeborenes Wissen angenommen werden. Eine zweite Annahme aus der generativistischen Tradition, gegen die sich die konnektionistische Forschung auflehnt, betrifft die behauptete Dichotomie von Kompetenz und Performanz. Mit dieser Dichotomie wird zum einen angenommen, dass es eine theoretisch ungebundene Kompetenz der Sprachverarbeitung gibt, die isoliert betrachtet nicht limitiert ist. Die tatsächlich beobachtbare Sprachverwendung weist nur deshalb Fehler und Beschränkungen auf, weil sie von einer durch unsere kognitiven Ressourcen limitierten Performanz umgesetzt wird. Zum anderen wird angenommen, dass alleine die Kompetenz unsere grammatische Kapazität beschreibt und die Kompetenz in ihren Ausprägungen unabhängig von der Performanz ist, d.h. der Sprachgebrauch nimmt keinen Einfluss auf die Entwicklung der Kompetenz. Da konnektionistische Modelle nicht über eine unlimitierte Kompetenz verfügen, sollten sie unter anderem nicht dazu in der Lage sein, die Produktivität der Sprachverarbeitung und -produktion abzubilden.

Zu den untersuchten Aspekten, die entgegen diesen Annahmen die prinzipielle Fähigkeit von SRNs zur Modellierung von Spracherwerb und Sprachverarbeitung belegen sollten, gehören:

- die Distinktion und Klassifizierung grundlegender syntaktischer Klassen (z.B. Nomen, Verben) und grammatischer Relationen (z.B. Subjekt-Objekt-Relation),
- die Produktivität der Sprache im Sinne einer potentiell infiniten Produktivität eines finiten Systems und in diesem Zusammenhang die Produktion und Verarbeitung (potentiell infiniten) rekursiver Strukturen (z.B. Mehrfach-Relativsatz-Einbettungen),

- Strukturen ohne direkt beobachtbare Evidenz wie zum Beispiel linguistische Gaps und Traces (etwa bei der Bildung von Fragesätzen mit Fragewort) sowie umgekehrt zufällige gegenüber intendierten Gaps,
- die syntaktische und semantische Systematizität der Verarbeitung, d.h. die Generalisierung von Wörtern und die Entwicklung von kontextunabhängigem Verwendungswissen (vgl. Kap. 3.1.4).

#### 3.1.3.1. Spracherwerbsfähigkeit

SRNs können syntaktische Klassen unterscheiden und darüber in Abhängigkeit vom sequenziellen Kontext korrekte Vorhersagen für Folgewörter leisten. In einer grundlegenden Studie untersucht Elman (1991) die Frage, wie Sprache und syntaktische Struktur in konnektionistischen Netzwerken repräsentiert werden: Welcher Natur sind linguistische Repräsentationen, wie werden komplexe strukturelle Relationen repräsentiert und wie kann die infinite Natur von Sprache in einem finiten System mit begrenzten Ressourcen erfasst werden? Hierfür testet er die Fähigkeit von SRNs, eine Reihe systematischer Eigenschaften sprachlicher Strukturen zu verarbeiten: Die Numerus-Kongruenz des Satzsubjekts zum finiten Verb, die Argumentstruktur verschiedener Verben (intransitive, transitive, ditransitive und optional transitive Verben), die Veränderung der Argument-Ordnung und der Abhängigkeiten durch Einbettung von Relativsätzen und Rekursion mit multipler Einbettungsstruktur. Anhand der Details der Vorhersagen auf kritischen Positionen wird demonstriert, dass SRNs immer korrekte Wortformen aktivieren und dabei die systematischen Eigenschaften erfüllen.

**Bedeutung initial reduzierter Komplexität** In der ersten Studie verwendet Elman (1991) einen Trainingsaufbau mit stufenweiser Erhöhung der Komplexität des Trainingskorpus über vier Schritte, in denen der relative Anteil von komplexen Satzstrukturen mit Relativsätzen zu einfachen Hauptsatz-Satzstrukturen erweitert wird. Dieses gestufte Training war eine Konsequenz schlechter Performanzdaten in Pilotstudien und führte zu einem schnellen und erfolgreichen Erwerb komplexer syntaktischer Strukturen. In einer Fortsetzung der Studie (Elman, 1993) wird dieser Aspekt des verbesserten Erwerbs komplexer Strukturen auf Basis stabiler Eigenschaften für einfache Strukturen gezielt untersucht. In der ersten Simulationsreihe wurde die Steigerung der Komplexität wieder über eine sukzessive Veränderung des Anteils komplexer Strukturen in den Trainings-Stimuli erzielt. In einer zweiten Simulationsreihe wurden die SRNs so manipuliert, dass der Kontext-Layer nach einer vorgegebenen Anzahl von Zeitschritten gelöscht und damit das dynamische Gedächtnis des Modells limitiert wurde. Zu Beginn erfolgte die Löschung nach kurzen Sequenzen (3-4 Zeitschritte). Mit fortschreitendem Training wurden die Sequenzen stufenweise erhöht und schließlich wurde das Netzwerk ohne Beschränkungen trainiert. Mit dieser Systematik wurde gezeigt, dass SRNs nicht nur robust gegenüber dieser initialen Beschränkung sind, sondern dass diese Beschränkung mitunter sogar eine Voraussetzung für den erfolgreichen Erwerb komplexer Strukturen ist. Die Beschränkung dient als Filter für die Komplexität des Inputs, da sie anfangs nur sehr kurze Sequenzen zulässt. Damit kann das Netzwerk stufenweise komplexere Zusammenhänge über größere Sequenzen auf der Basis von Regularitäten in einfacheren Strukturen erwerben. Aus den Ergebnissen

von Elman (1993) wurden Konsequenzen für die Bedeutung der langsamen Reifung des kognitiven Systems in der frühen Entwicklungsphase diskutiert und darin unter anderem Erklärungen für kritische Perioden des Spracherwerbs<sup>5</sup> gesucht.

Warum erfordert der erfolgreiche Erwerb komplexer Strukturen in manchen Fällen den Beginn des Trainings mit voller Komplexität, während in anderen Fällen nur eine allmähliche Erhöhung der Komplexität zum Erfolg führt? Elman (1993) leitet über diese Fragestellung eine Reihe grundlegender Annahmen über statistisches Lernen in konnektionistischen Netzen ab und bietet daraus einen Erklärungsansatz: Konnektionistische Netze leisten nicht die Ableitung von Kookkurrenz-Statistiken (statistische Induktion als Ergebnis des Lernens), sondern Näherungen für die Funktionen, die den Daten zugrunde liegen (statistische Induktion als treibende Kraft des Lernens). Dabei sind sie sensitiv für die statistische Struktur der Daten. Diese Sensitivität führt dazu, dass Netzwerke von der Repräsentativität des Trainings-Sets abhängig sind, denn je kleiner das Sample ist, desto größer ist die Gefahr, dass keine geeignete Funktion zur Beschreibung der Daten abgeleitet werden kann. Hypothesen für diese Funktion werden in den Gewichten gespeichert. Das Lernen anhand des Gradientenabstiegsverfahrens führt dazu, dass bei der Anpassung der Hypothesen nicht gänzlich verschiedene Hypothesen getestet werden, sondern bestehende Hypothesen sukzessive und geringfügig verändert werden. Dabei ist das Gradientenabstiegsverfahren anfällig für lokale Minima – und damit für eine Fixierung einer falschen Hypothese. Zudem erhalten wir bei Verwendung einer sigmoiden Aktivierungsfunktion und einer mittleren Randomisierung der initialen Input-Gewichte eine stärkere Anpassung der Gewichte in frühen Phasen des Trainings, da diese Funktion in mittleren Aktivierungsbereichen die stärkste Steigung aufweist. Anhand dieser Merkmale erklärt Elman das Verhalten der SRNs in seinen Simulationen: Enthält das Datenset von Anfang an komplexe Strukturen, ist die Gefahr der Fixierung falscher Hypothesen größer, als wenn anfangs nur Hypothesen für einfache Strukturen gebildet werden, aus denen dann Hypothesen für die komplexen Strukturen angepasst werden können.

Allerdings legen Rohde und Plaut (1999) eine Serie von Simulationsstudien mit SRNs dar, in denen sie die Ergebnisse von Elman (1993) kritisch überprüfen und deren Befunde Elmans Annahmen eines positiven oder sogar notwendigen Effekts der anfänglichen Limitierung der Stimuluskomplexität oder der Gedächtniskapazität widersprechen. Vielmehr zeigen sie sogar, dass die Simplifizierung der Stimuli hinderlich für das Lernen syntaktischer Strukturen ist, vor allem dann, wenn die Sprache komplexer und damit realistischer gemacht wird wie etwa durch die Einführung semantischer Constraints. Bei weiteren Variationen stellen sie zudem fest, dass die Stärke bei der initialen Randomisierung der Gewichte ein bedeutsamer Faktor ist: Durch sie kann manipuliert werden, wie früh verschiedene Lernplateaus durchbrochen werden – je größer der Wertebereich für die Initialisierung, desto früher erfolgt der Lernfortschritt. Ebenfalls getestet wurde die Manipulation der SRNs mit initialer Gedächtnisbeschränkung und sukzessiver Kapazitätserweiterung. Zwar brachten diese Versionen bessere Ergebnisse als die Versionen mit anfänglich reduzierten Stimuli. Allerdings zeigte sich kein Unterschied gegenüber der Performanz mit voller

---

<sup>5</sup>Aus Einzelfallstudien ist unter anderem bekannt, dass der Erwerb vor allem von komplexeren grammatischen Strukturen stark beeinträchtigt ist, wenn die Sprache erst nach mehreren Jahren sprachlicher Deprivation erworben wird. Zudem wird beim Spracherwerb zwischen frühen gegenüber späten Lernern unterschieden, die in den frühen Phasen des Erwerbs vergleichsweise wenig produktiv sind, woraus sich in der Folge Nachteile in der Sprachfähigkeit ergeben (vgl. Szagun, 2000).

Komplexität und unbeschränktem Gedächtnis. Die Manipulation der Gedächtnisspanne war also einerseits nicht hinderlich für den Erwerb, andererseits brachte sie auch keine Vorteile mit sich.

Insgesamt zeigt sich in den Simulationen von Rohde und Plaut (1999) eine gute Performanz für den Erwerb komplexer syntaktischer Strukturen. Die Autoren weisen diese Lern- und Generalisierungsfähigkeit der Nutzung impliziter negativer Evidenz<sup>6</sup> zu, die durch die Implementierung der Vorhersageaufgabe in SRNs zustandekommt. Während sie also ein positives Ergebnis für die allgemeine Spracherwerbskompetenz von SRNs erzielen, lehnen sie die Annahme ab, dass erfolgreiches Lernen in dieser Architektur irgendwelcher Reifungsmechanismen bedarf. Als Erklärung für das erfolgreiche Lernen trotz vollständigem Input bieten sie eine alternative Sichtweise auf interne Komplexitätsreduktion an: SRNs beginnen ganz von selbst mit einem geringen Gedächtnis, da die internen Repräsentationen, die das Netzwerk für den Erwerb jeweils komplexerer Relationen benötigt, erst für einfachere Relationen entwickelt werden muss. Anders formuliert: Die Netzwerke tendieren dazu, erst einfache Relationen zu lernen, da sie davor nicht die repräsentationale Kapazität besitzen, um mit komplexeren Relationen umzugehen. Der gegensätzliche Befund von Rohde und Plaut (1999) hinsichtlich der expliziten Steuerung der Komplexitätsreduktion weist damit die Ergebnisse von Elman (1993) als Artefakt der Konfiguration der Trainingsparameter zurück.

**Nutzung statistischer Informationen** J. Lewis und Elman (2001a, 2001b) stellen eine weitere konnektionistische Untersuchung vor, in der ein Poverty-of-the-Stimulus-Argument überprüft wird. Sie modellieren den Erwerb der Struktur von Relativsätzen und die Generalisierung der Anwendung dieser Struktur auf die Subjektposition in Auxiliarfällen. Zur Bildung dieser Satzstruktur muss eine Inversion von Subjekt und Verb erfolgen. Diese Inversion erfordert allerdings eine strukturbasierte Generalisierung – eine lineare Generalisierung hingegen erzeugt eine falsche Äußerung. Dies wird in Beispiel 2 demonstriert: Während die Inversion des Satzes nur eine lineare Generalisierung erfordert (Verschiebung des ersten Verbs, Bsp. 2a), würde dieselbe Generalisierung bei Sätzen mit eingebettetem Relativsatz (Bsp. 2b) eine falsche Äußerung produzieren (Bsp. 2c). Nur eine strukturbasierte Generalisierung (Verschiebung des Matrixsatz-Verbs) produziert eine korrekte Äußerung (Bsp. 2d).

- (2) The boy is crazy.
  - a. Is the boy crazy?
  - b. The boy who is smoking is crazy.
  - c. \*Is the boy who smoking is crazy?
  - d. Is the boy who is smoking crazy?

Während solche komplexen Äußerungen in der kindgerichteten Sprache (KGS) praktisch nicht vorhanden sind, werden falsche Äußerungen wie Satz 2c nie gebildet. Für solche

---

<sup>6</sup>Durch die ständige umweltgegebene Korrektur der (mitunter falschen) Vorhersagen durch positive Beispiele erfährt das Netzwerk implizit negative Evidenzen – falsche Vorhersagen, die nie bestätigt und deshalb beim Verstärkungslernen in SRNs fortschreitend geschwächt werden.

strukturbasierten Generalisierungen wird deshalb angenommen, dass der Input für ihren Erwerb nicht genügt und nur auf Grundlage angeborenen Wissens erworben werden kann (vgl. Chomsky, 1975). Wie J. Lewis und Elman (2001a) aber zeigen können, generieren SRNs für solche Strukturen korrekte Vorhersagen auf der kritischen Position des finiten Verbs des Relativsatzes. Für eine weitere Simulation zu diesem Problem haben die Autoren ein Korpus generiert, das die Merkmale von KGS abbildet.<sup>7</sup> Auch hierfür wurden korrekte Vorhersagen auf den kritischen Positionen bestätigt. Die Autoren begründen diese Fähigkeit über die Annahme, dass SRNs statistische Eigenschaften des Inputs nutzen können, über die die fehlende negative Evidenz durch indirekte positive Evidenz ersetzt wird. Reali und Christiansen (2005) präsentieren weitere Untersuchungen zu diesem Gegenstand anhand von Korpusanalysen mit Bigramm- und Trigramm-Modellen<sup>8</sup> sowie mit erweiterten SRN-Simulationen, die diese Ergebnisse bestätigen und die Bedeutung versteckter statistischer Informationen hervorheben.

Die Erfassung versteckter statistischer Informationen spielt auch bei der Erkennung von Wortgrenzen im frühen Spracherwerb eine wichtige Rolle. Wir haben bereits die Erfassung von Wortgrenzen im Lautstrom (Elman, 1990) als Positionen mit hoher Entropie (Umschreibung nach Rohde, 2002) erwähnt. Allerdings müssen die Wortgrenzen aus der Fehlerkurve von außen als solche interpretiert werden und die Ergebnisse beruhen auf sehr reduzierten Stimuli. In einer Reihe späterer Studien wurde mit SRNs die Lautstromsegmentierung hinsichtlich weiterer kritischer Aspekte und mit realistischeren Stimuli (basierend auf einem KGS-Korpus) untersucht (Christiansen, Allen & Seidenberg, 1998; Reali, Christiansen & Monaghan, 2003; Christiansen, Conway & Curtin, 2005). Im Zentrum stand dabei die Fähigkeit von SRNs, multiple Cues aus unterschiedlichen Quellen zu integrieren und aus den kombinierten statistischen Merkmalen versteckte Cues abzuleiten, die an der Oberfläche der Verteilungen im Input nicht erkennbar sind. Hierzu wurden neben unterscheidbaren Phonemen mit zusätzlichen Units Äußerungsgrenzen (Ende einer längeren Wortsequenz) sowie Betonungsmerkmale (auf den zwei Betonungsebenen primary vs. secondary stress) kodiert und koaktiviert – Informationen, die im Sprachinput natürlich gegeben sind. Es wurde erwartet, dass die Aktivierung der Äußerungsgrenze als Markierung von Wortgrenzen generalisiert wird. Während die einzelnen Informationen keine hinreichenden Informationen für die Erfassung von Wortgrenzen bieten<sup>9</sup>, ermöglichte die Darbietung aller drei Informationsquellen die erwartete Generalisierung mit hoher Genauigkeit. Die hierfür genutzten statistischen Informationen zweiter Ordnung – versteckte Informationen, die aus der Konjunktion statistischer Merkmale aus verschiedenen Quellen abgeleitet werden müssen – und die Rolle der Integration verschiedener Aufgaben (primäre, abgeleitete und unmittelbare Aufgaben) werden in Christiansen et al. (2005) diskutiert.

---

<sup>7</sup>Als Grundlage für die vorhandenen Strukturen und Frequenzen wurde das KGS-Korpus des CHILDES-Projekts (MacWhinney, 2000) verwendet.

<sup>8</sup>Ein Ngramm ist eine Folge von n Elementen (z.B. Buchstaben oder Wörter). In Ngramm-Modellen wird korpusgestützt die statistische Wahrscheinlichkeit für die Folge eines bestimmten Elements nach einem Fenster von n Elementen berechnet. So wird in einem Trigramm-Modell für Wörter in einem Korpus die Wahrscheinlichkeit der Wörter in einem Kontext von drei Wörtern berechnet.

<sup>9</sup>Auch der Verlauf des Fehlers, für den in der Studie von Elman (1990) ein Anstieg an Wortgrenzen gezeigt wurde, bot für dieses realistische Korpus keine verlässlichen Hinweise auf Wortgrenzen, sondern allenfalls für höherfrequente Silbenstrukturen.

#### 3.1.3.2. Interne Repräsentation struktureller Abhängigkeiten

SRNs können also rein induktiv grammatische Zusammenhänge und Regelmäßigkeiten identifizieren und umsetzen. Welche Eigenschaften ermöglichen der Architektur diese Leistungsfähigkeit? Oder anders gefragt: Was muss das Netzwerk erwerben und welche internen Strukturen muss es dabei entwickeln, um diese Performanz zu zeigen?

**Frequenz und Konkurrenz** Eine zentrale Rolle nimmt die Kategorisierung des Inputs ein, die aus der Struktur der Gewichtsmatrizen zwischen den Layern und der Struktur der Aktivierungsmuster des Hidden-Layers abgeleitet und nachvollzogen werden kann. Die Kategorisierung der Wörter erfolgt bei Anwendung der Wortvorhersage-Aufgabe über die Eigenschaften ihrer Frequenz und ihrer Verteilung im Verhältnis zu den anderen Wörtern. Ähnlich wie in kollokationsbasierten Ansätzen semantischer Repräsentation – etwa das Hyperspace Analogue of Language (HAL, Burgess & Lund, 1997), das Correlated Occurrence Analogue to Lexical Semantics (COALS, Rohde, Gonnerman & Plaut, 2005) oder die Latent Semantic Analysis (LSA, Landauer, Foltz & Laham, 1998)<sup>10</sup> – wird ein Wort anhand seiner möglichen syntaktischen Kontexte klassifiziert. Diese kontextuellen Zusammenhänge müssen im Verlauf des Trainings auf die interne Struktur des Modells abgebildet werden.

Wir sind bereits in Kapitel 2.3 kurz darauf zu sprechen gekommen, welche Perspektiven die internen Repräsentationen auf die Interpretation konzeptuellen Wissens in konnektionistischen Netzwerken bieten. Im Zentrum steht die Ähnlichkeitsstruktur im n-dimensionalen Repräsentationsraum für Aktivierungsmuster der Hidden-Layer und Gewichtsstrukturen zu Input- oder Output-Units mit distinktiver Inhaltsrepräsentation. Die Distanzen zwischen Verortungen in der Repräsentationsstruktur sowie Gruppierungen und Gruppierungsunterschiede entsprechen der abgebildeten konzeptuellen Struktur und dem Kategorisierungswissen des Netzwerks. Die interne Repräsentationsstruktur, die sich in SRNs für die Wortvorhersage-Aufgabe entwickelt, dient der Erfassung und Abbildung von Konkurrenzstrukturen und ermöglicht die Generierung der korrekten Verteilung der Wortfolgewardrscheinlichkeit (vgl. Elman, 1990). Dabei entspricht die abgebildete Konkurrenzstruktur zunächst einmal den syntaktischen Wortklassen, die den Strukturen des Trainingskorpus zugrunde liegen. Hier sei nochmals darauf hingewiesen, dass diese Wortklassen dem Netzwerk niemals explizit vermittelt wurden. Das Netzwerk hat die Klassifizierung verschiedener Wortklassen (v.a. Verben und Nomen) allein aus der Verteilungsstruktur des Inputs abgeleitet und selbstorganisierend auf die interne Struktur abgebildet.

---

<sup>10</sup>Latent Semantic Analysis (Landauer et al., 1998) ist eine Methode zur Extraktion der kontextuellen gebrauchsbasierten Bedeutung von Wörtern anhand statistischer computergestützter Verfahren (als statistisches Verfahren entspricht die LSA einer generalisierten Faktorenanalyse). Die Ähnlichkeit des Kontextes, in dem ein Wort auftauchen kann, bestimmt die Ähnlichkeit seiner Bedeutung zu anderen Wörtern und Wortgruppen. LSA produziert anhand linguistischer Korpora Ähnlichkeitsmaße für Wort-Wort-, Wort-Passage- sowie Passage-Passage-Relationen. Für so extrahierte Ähnlichkeitsmaße wurde eine hohe Übereinstimmung zu empirischen Befunden in verschiedenen Domänen festgestellt, unter anderem für Wortsortieraufgaben, Kategorienbeurteilungen, lexikalischen Priming-Tests, Einschätzungen der Bewertung von Übergangskohärenzen in Texten und Bewertungen von Qualität und Quantität des in Essays enthaltenen Wissens.

Die geleistete Klassifizierung, die von Elman (1990) demonstriert wird, geht aber weit über diese Grobstruktur hinaus. So zeigen sich in Clusteranalysen auch weitergehende strukturelle und semantische Distinktionen: Für die Verben zeigt sich eine Unterscheidung intransitiver, transitiver und ditransitiver Verben, die Nomen unterscheiden sich nach den Merkmalen belebt vs. unbelebt, und innerhalb dieser Feinstrukturen finden sich weitere semantische Distinktionen, zum Beispiel zwischen Mensch und Tier sowie zwischen essbaren und zerbrechlichen Materialien. Dieses Ergebnis bietet Anlass, die Annahme abzulehnen, dass für den Erwerb grundlegender syntaktischer Klassen angeborene Mechanismen benötigt werden, die die Erfassung der Distinktionen ihrer grammatischen und semantischen Verwendung ermöglichen.

**Kontextabhängige Strukturierung** Mit der Erfassung von kategorialen Abhängigkeiten ist die Abbildungsmächtigkeit aber noch lange nicht ausgeschöpft. So zeigte sich in einer weiteren Clusteranalyse, in der für jedes im Datenset enthaltene Wort der syntaktische Kontext als weiteres Unterscheidungsmerkmal mitaufgeführt wurde, dass in der internen Struktur auch feine Unterschiede des Kontexts erfasst werden. Damit werden Fragen der Kontextsensitivität der Verarbeitung, der hierarchischen kategorialen Struktur und der Unterscheidung von spezifischen Ereignissen (Tokens) und übergeordneten Ereigniskategorien (Types) behandelt. Types und Tokens werden in symbolischen Systemen zumeist explizit unterschieden und erfordern häufig Indexierungs- oder Bindungsoperationen – in diesem Modell können die Beziehungen innerhalb eines Systems ohne Annahme zusätzlicher Operationen erfasst werden (Elman, 1990, S. 23). Hier sei nochmals daran erinnert, dass es bei der Derivation der statistischen Struktur des Inputs nicht nur um deren Abbildung im Sinne einer Wahrscheinlichkeitstabelle geht. Vielmehr erfassen SRNs geleitet durch statistische Inferenzen die Funktionen, die den Verteilungen und Zusammenhängen des Inputs zugrunde liegen.

In der Studie von Elman (1991) stehen weitere Fragen zur Repräsentation von Sprache im Konnektionismus im Vordergrund: Was ist die Natur linguistischer Repräsentationen in konnektionistischen Netzwerken, wie werden komplexe strukturelle Relationen repräsentiert und wie kann in diesem finiten System die infinite Natur von Sprache erfasst werden? Elman demonstriert, wie größere grammatische Strukturen einschließlich kompletter komplexer Sätze mit multiplen Einbettungen als Trajektorien im hochdimensionalen Zustandsraum erfasst werden können. Anhand einer Hauptkomponentenanalyse können die Zusammenhänge in den Aktivierungsmustern des Hidden-Layers als Verortungen im und Bewegungsmuster durch den Zustandsraum visualisiert werden. Diese Bewegungsmuster zeigen spezifische Strukturen und Unterschiede für verschiedene syntaktische Strukturen, in denen sich die Dynamik des Netzwerks bei der Verarbeitung von Wortsequenzen widerspiegelt. Im Sinne von wohldefinierten begehbaren Pfaden lässt sich diese Dynamik als Enkodierung des grammatischen Wissens in den Netzwerken deuten (Elman, 1991, S. 210). Dabei werden in verschiedenen Komponenten im Zustandsraum minimale Differenzen innerhalb der syntaktischen Pfade ersichtlich, die grammatische Abhängigkeiten wie etwa Subjekt-Prädikat-Kongruenz etablieren. In Beziehung zur Konstruktionsgrammatik beschreibt Elman (2004) diese Trajektorien als syntaktische Konstruktionen, in denen sich die Relation zwischen Lexikon und Grammatik manifestiert.

Im Laufe der Derivation der kategorialen Struktur des Inputs verbessert sich auch sprunghaft die Generalisierungsleistung des Netzwerks. Dieses Verhalten bietet einen Erklärungsansatz für das Erreichen einer kritischen Masse im Worterwerb und den dort auftretenden plötzlichen Vokabelspurt im kindlichen Spracherwerb, der sich empirisch zumeist ab etwa 1;8 Jahren feststellen lässt. Diesen Zusammenhang legen auch weitere Simulationsstudien von (Borovsky & Elman, 2006) nahe, in denen sie Inputmenge, Satzkomplexität und Verteilung lexikalischer wie kategorialer Frequenzen manipulieren. Zur Beurteilung der Erwerbsqualität ziehen sie eine Kalkulation der globalen Kohärenz in der Strukturierung der Hidden-Layer auf Basis vergleichender hierarchischer Clusteranalysen heran (Average Precision). Die Resultate zeigen eine starke Korrelation zwischen der Verbesserung der kategorialen Struktur und der weiteren Worterwerbsfähigkeit. Die Simulationen legen eine erfahrungsabhängige Veränderung kognitiver Fähigkeiten als Erklärung interindividueller Erwerbsverläufe und Lernerunterschiede nahe.

Die interne Repräsentation und Dynamik von Sprache in SRNs bietet auch eine ganz neue Perspektive auf Sprachverarbeitung: Es wird kein eigener lexikalischer Zugriffsprozess von der weiteren syntaktischen Verarbeitung unterschieden. Darüberhinaus gibt es auch keine Repräsentation von isolierten Wörtern mehr, sondern nur noch kontextgebundene Wortzustände, die sich in ihren Eigenschaften in der Auswirkung auf die Dynamik im Netzwerk unterscheiden. Gerade dieser Unterschied wird in späteren Veröffentlichungen verstärkt diskutiert (Elman, 2004, 2009), denn er ist eine klare Absage an ein bis dato allgemein akzeptiertes Konstrukt der Beschreibung menschlicher Sprachfähigkeit – das Mentale Lexikon.<sup>11</sup> In Elmans SRN-basierter Gegenkonzeption zum Mentalen Lexikon als Speicher von Form-Bedeutung-Einträgen werden Wörter nicht als passive, sondern als aktive Einheiten – als Operatoren – konzipiert. Dies impliziert auch eine neue Interpretation für das Konzept Wortbedeutung (Elman, 2004, S. 306): „The fundamental suggestion of the present proposal is to treat words as stimuli, whose ‚meaning‘ [Hervorhebung im Original] lies in the causal effects they have on mental states.“

Diese Abkehr von der Speichermetapher des Mentalen Lexikons trägt eine Lösung für ein zentrales Problem dieses Konstrukts in sich: Die Frage, welchen Umfang ein lexikalischer Eintrag besitzt und welche Informationsstrukturen gespeichert und beim Abruf aktiviert werden. In einem Review empirischer Daten stellt Elman (2009) klassischen Theorien des Mentalen Lexikons Belege einer umfangreichen abgerufenen Informationsstruktur<sup>12</sup> entgegen, deren Reichhaltigkeit in diesen Ansätzen kaum sinnvoll erfasst werden kann. In Elmans konnektionistischer Gegenkonzeption erfolgt die unmittelbare Aktivierung des relevanten Wissens automatisch und ohne spezifizierte Speicherung. Die Ad-hoc-Kontextualisierung der fast immer kontextabhängigen Wortbedeutung erfolgt ebenfalls als automatischer Prozess.

---

<sup>11</sup>Auf ähnlicher Grundlage stellt auch Bullinaria (1995) die provokante Frage „Modelling Lexical Decision: Who needs a lexicon?“ (Titel der Studie). Die Studie behandelt die Modellierung von qualitativen Effekten und Reaktionszeiten in lexikalischen Entscheidungsaufgaben mit einfachen konnektionistischen Netzwerken (FFNs).

<sup>12</sup>Eine lexikalische Repräsentation müsste unter anderem enthalten: verbsspezifische syntaktisch-strukturelle Erwartungen, präferierte thematische Rollen bzw. Argumentstruktur, Spezifizierungen durch den Aspekt des Verbs, durch Interaktionen der Argumentstruktur, durch das Instrument der Handlung und durch den situationalen Kontext.

### 3.1.3.3. Exkurs: Analyse interner Repräsentationen

Die Untersuchung der internen Strukturierung von Netzwerken spielt eine wichtige Rolle in konnektionistischen Modellierungen: Sie bietet einen entscheidenden analytischen Zugang zu den Lösungen, die in den Modellen selbstorganisiert entwickelt werden. Deshalb wollen wir die gängigsten Verfahren in diesem Exkurs kurz vorstellen. Zur Untersuchung werden mathematische Distanz- bzw. Ähnlichkeitsmaße und darauf aufbauend strukturaufdeckende multivariate statistische Analysemethoden herangezogen. Wie wir schon erläutert haben (vgl. Kap. 2.2.2), wird dazu die Aktivierung eines Layers als Vektor aufgefasst, dessen Stellen durch die Units im Layer repräsentiert werden. Die Aktivierungen der einzelnen Output-Units werden als Wert der jeweiligen Vektorstelle erfasst. Jeder Vektor zeigt auf einen Punkt, der einem repräsentationalen Zustand entspricht, der in einem hochdimensionalen Zustandsraum verortet ist. In diesem Raum nehmen die repräsentationalen Zustände Positionen ein, deren Verortung verglichen und als Distanz erfasst werden kann. Es gibt aber auch spezifischere Maße, die von dieser Vektor-Betrachtungsweise abweichen und Aktivierungsstärken einzelner Units und Unit-Gruppen heranziehen bzw. Verhältnisse der Aktivierungsstärken zwischen Units und Unit-Gruppen vergleichen.

Zur Analyse der internen Repräsentationen vergleicht man die internen Zustände in Abhängigkeit ihrer korrespondierenden Input- und Output-Zustände. Dies können Vektoren der Hidden-Layer sein, auf denen für die verschiedenen Stimuli spezifische Aktivierungsmuster gebildet werden, die sich im Netzwerk selbstorganisiert zur Lösung der Abbildung von Input- auf Target-Stimuli entwickelt haben. Gleichmaßen können aber auch die Gewichtsvektoren zwischen den Schichten in Abhängigkeit von der sendenden oder der empfangenden Unit herangezogen werden.<sup>13</sup> Die Beziehungen zwischen verschiedenen Stimuli (Input- und Output-Muster) werden anhand ihrer Distanz innerhalb des Zustandsraums erfasst. Während die Gewichtsmatrix nur eine absolute Verortung im Repräsentationsraum ermöglicht, da die Gewichte beim Testen fixiert und damit unabhängig vom aktuellen Input sind, ermöglicht die Analyse der Aktivierungsmuster des Hidden-Layers auch die Unterscheidung von Input-Mustern in Abhängigkeit vom aktuellen sequenziellen Kontext. Zudem können damit komplexe Ereignisse, etwa hierarchisch-sequenzielle Input-Strukturen (wie z.B. natürlichsprachliche Sätze), als Trajektorien durch den Zustandsraum beschrieben werden.

Um die Gliederung dieses internen Raums und die systematischen Zusammenhänge der internen Repräsentationen für verschiedene Zustände zu erfassen und abzubilden, verwendet man verschiedene statistische Verfahren aus dem Instrumentarium der multivariaten strukturentdeckenden Verfahren, die diesen Zustandsraum reduzieren und die Zusammenhänge, Hierarchien und Ähnlichkeitsrepräsentationen in kompakter aussagekräftiger Weise abbilden. Dabei kann man grob zwischen clusteranalytischen und dimensionsreduzierenden Verfahren unterscheiden. Verwendet werden vor allem die Hierarchische

---

<sup>13</sup>Wissen ist verteilt über das gesamte Netzwerk in den Gewichten gespeichert, die für Inputsignale distinktive Muster der Aktivierungspropagation erzeugen. Bei der Analyse von Gewichtsmatrizen werden vor allem die vom Input-Layer wegführenden und die zum Output-Layer hinführenden Gewichte relevant. Dies setzt in der Regel eine lokalistische Stimulus-Kodierung voraus, über die dem Gewichtsvektor einer Unit (d.h. dem Vektor der Stärke aller Gewichte dieser Unit) ein sachlicher Inhalt zugewiesen werden kann. Die Gewichtsvektoren für verschiedene Stimuli lassen sich nun untereinander vergleichen (vgl. z.B. McClelland et al., 1989).

Clusteranalyse, die Hauptkomponentenanalyse sowie die Multidimensionale Skalierung. Neben diesen Verfahren finden sich weitere Methoden wie etwa die Sensitivitätsanalyse oder die Diskriminanzanalyse, mit denen versucht wird, den Beitrag kleinerer Einheiten des internen Repräsentationsraums zur Performanz des Netzwerks zu bestimmen. Wir werden die relevanten Verfahren nur kurz umreißen. Nähere Informationen finden sich in der einschlägigen Literatur (z.B. Bortz, 1999; StatSoft, 2010; Roessler & Ungerer, 2010; Baayen, 2008, für Umsetzungen in R), die auch den hier gezeigten Darstellungen zugrunde liegt. Einen Überblick über die verschiedenen in der konnektionistischen Literatur verwendeten Verfahren mit besonderem Blick auf verschiedene Nachteile und Beschränkungen ihrer Aufdeckungsmächtigkeit bietet zum Beispiel Bullinaria (1997).

**Clusteranalytische Verfahren** Bei der Clusteranalyse werden im Datensatz inhaltlich zusammenhängende Gruppen oder Klassen identifiziert, indem die Datenmenge durch Zusammenfassen von ähnlichen Datenpunkten zu Untergruppen reduziert wird. Für die Clusterung der Daten existieren verschiedene Verfahren: Hierarchisches Clustern, der k-means-Algorithmus und Fuzzy Clustering (für eine kurze Übersicht vgl. Wiedenbeck & Züll, 2001). Die Clusterung erfolgt über Distanzmaße (in Abhängigkeit des Skalenniveaus der abhängigen Variablen, für metrische Skalen etwa die euklidische Distanz), anhand derer man die Ähnlichkeit von Datenpunkten erfassen und ihre Zusammenfassung zu Clustern bestimmen kann. Die dabei entstehende Clusterstruktur kann über verschiedene Mittel visualisiert werden, gängig ist die Darstellung als Dendrogramm. Bei der hierarchischen Clusteranalyse wird zwischen agglomerativen und divisiven Verfahren unterschieden: Bei den agglomerativen Verfahren erfolgt die Clusterung von der kleinsten zur größten Gruppierung (bottom-up), bei den divisiven Verfahren von der größten zur kleinsten Gruppierung (top-down). Die agglomerativen Verfahren unterscheiden sich dabei in der Bestimmung der Gruppierungsauswahl ab der zweiten Ebene – d.h. sobald nicht mehr einzelne Datenpunkte, sondern Cluster in größere Cluster zusammengefasst werden sollen.<sup>14</sup>

In der konnektionistischen Forschung verwendet man in der Regel die hierarchische Clusteranalyse (vgl. z.B. Elman, 1990; Elman et al., 1996; Elman, 2004). Die gefundenen Clusterlösungen bieten Hinweise auf die Organisation des Repräsentationsraums, die das Netzwerk während den Trainingsphasen entwickelt hat. Bei der Umsetzung zur Analyse der internen Repräsentationen werden die einzelnen Units eines Layers als Variablen aufgefasst und ihr Aktivierungswert in Abhängigkeit der unterscheidbaren Inputmuster erfasst. Diese Aktivierungsmuster betrachtet man als Vektoren und berechnet ihre Distanz zueinander. Damit ergibt sich eine Distanzmatrix für die Vektoren. Die Distanz der Vektoren für die Inputmuster entspricht der Unterschiedlichkeit der Repräsentation der Inputmuster im Vektorraum. Um die Distanzmatrix erstellen zu können, müssen alle Inputmuster distinktiv sein. Zur Verwendung für SNRs mit Wortvorhersage-Aufgabe müssen die Daten zumeist aufbereitet werden, da in den Korpora Wörter zumeist mehrfach erscheinen. Dies kann zum Beispiel auf folgende Weisen geschehen: (a) Gleiche Wörter erhalten ein

---

<sup>14</sup>Agglomerative Verfahren in Auswahl: Das Single Linkage-Verfahren bestimmt die Zusammenfassung über die Distanz der Vertreter mit dem geringsten Abstand, das Complete Linkage-Verfahren bestimmt die Zusammenfassung über die Distanz der Vertreter mit dem größten Abstand, das Average Linkage-Verfahren bestimmt die Zusammenfassung über die Distanz der Mittelwerte der Vertreter der Cluster. Über die Auswahl des Verfahrens können Eigenschaften der Clusterung manipuliert werden, so etwa die Gruppengröße der Cluster.

künstliches Unterscheidungsmerkmal, etwa eine Numerierung, (b) für gleiche Wörter wird der Mittelwert der Vektorpositionen erfasst, über den die jeweils gleichen Wörter einen gemeinsamen Stellvertreter erhalten, (c) für alle Wörter werden weitere unterscheidende Merkmale wie etwa der syntaktische Kontext oder ihre Kollokation erfasst.<sup>15</sup>

Zur Interpretation der Clusterlösung müssen externe Vergleiche mit bekannten Gruppierungen herangezogen werden, etwa theoretische syntaktische oder semantische Klassen. Zur Bestimmung der Signifikanz (bzw. als Annäherung daran) kann man zum einen die Reproduzierbarkeit der Lösung für Teilmengen des Datensatzes testen<sup>16</sup>, zum anderen kann man die gefundene Lösung mit Lösungen auf der Basis anderer Algorithmen oder alternativer (etwa faktorenanalytischer) Verfahren vergleichen. Nachteile der hierarchischen Clusteranalyse sind unter anderem die feste Zuordnung von Datenpunkten zu nur einem Cluster, auch wenn sie weiteren Gruppierungen zugeordnet werden könnten, sowie die mitunter verzerrte Zuordnung von Datenpunkten durch die Neubestimmung der Cluster-Verortung, die vom jeweils verwendeten Algorithmus abhängt. Diese Nachteile legen eine zusätzliche Überprüfung der Strukturlösung mit alternativen Verfahren nahe, die die relativen Distanzen und Verortungen im Vektorraum besser aufrechterhalten (vgl. Bullinaria, 1997; R. Frank et al., 2005).

**Dimensionsreduzierende Verfahren** Bei den dimensionsreduzierenden Verfahren erfolgt die Reduktion der Datenmenge bzw. der Dimensionen des Datensatzes über die Identifikation zugrundeliegender unabhängiger Faktoren oder Hauptkomponenten, über die der Datensatz beschrieben werden kann. Die Ermittlung dieser Faktoren oder Komponenten kann durch verschiedene statistische Verfahren und Algorithmen erreicht werden, beispielsweise die explorative Faktorenanalyse, die Hauptkomponentenanalyse oder die Korrespondenzanalyse. Ziel ist es, die Eigenschaften des Datensatzes durch die dimensionale Reduktion visualisierbar zu machen. Durch die Abbildung in einem zwei- oder dreidimensionalen Koordinatensystem lassen sich Kategorisierungsmerkmale als räumliche Zusammenhänge und Unterschiede zwischen Datenpunkten identifizieren. Häufig wird für die Analyse interner Repräsentationen konnektionistischer Netzwerke die Hauptkomponentenanalyse eingesetzt (vgl. z.B. Elman, 1991, 1993). Als Alternative zur Faktorenanalyse wird häufig die Multidimensionale Skalierung verwendet (z.B. Botvinick & Plaut, 2004), die allerdings auf etwas anderen Prinzipien basiert.

Bei der Umsetzung der Analyse interner Repräsentationen werden die einzelnen Units eines Layers als unabhängige Variablen aufgefasst. Die Beobachtungen für diese Variablen sind die Aktivierungswerte (als unabhängige Variable) in Abhängigkeit des aktuellen Inputmusters. Jeder Variablen (d.h. jeder Unit) wird ein Anteil an der Repräsentation des Inputs zugeschrieben. Für die Variablen wird nun eine Korrelationsmatrix erstellt. Bei der explorativen Faktorenanalyse und der Hauptkomponentenanalyse werden die

---

<sup>15</sup>Alle Varianten haben ihre Vor- und Nachteile: Während die erste Variante zumeist zu sehr unübersichtlichen Darstellungen im Dendrogramm führt, kann die Lösung über Mittelwerte zu künstlichen Lösungen führen, da Information über möglicherweise sehr unterschiedliche Vertreter verloren geht. Bei der Aufnahme zusätzlicher Unterscheidungsmerkmale nimmt die Auswahl dieser Merkmale natürlich Einfluss auf die Clusterstruktur.

<sup>16</sup>Ein solches Sampling der Daten erfolgt zum Beispiel in der Consensus-Analyse (vgl. Baayen, 2008, S. 159).

zugrundeliegende Faktoren über hochkorrelierte Variablen Gruppen gebildet, deren Vertreter gegenüber den Vertretern anderer Faktoren möglichst gering korreliert sind. In beiden Fällen liegt die Annahme zugrunde, dass die Daten durch ein lineares Modell beschrieben werden können. Eine Alternative zur Faktorenanalyse ist die Multidimensionale Skalierung, bei der der Datensatz auf eine Weise in einen niedrig-dimensionalen Raum transformiert wird, die die relativen Distanzen der Daten so gut wie möglich erhält (eine kurze Übersicht bietet Borg, 2000). Bullinaria (1997, S. 11) erwähnt zwei Probleme, die sowohl für die Hauptkomponentenanalyse als auch für die Multidimensionale Skalierung gelten: Zum einen unterscheiden sie nicht zwischen der unterschiedlich gewichteten Bedeutung der verschiedenen Units für die Performanz (alle Units werden als gleich bedeutsam erachtet, obwohl sie einen unterschiedlich großen Anteil an den Ausprägungen der Performanz haben können), zum anderen können kaum alle relevanten Aspekte der internen Repräsentation in nur zwei Dimensionen erfasst werden.

**Verfahren mit partieller Analyse** Neben den strukturaufdeckenden multivariaten Verfahren werden weitere Verfahren angewandt, deren Grundprinzip sich damit zusammenfassen lässt, dass sie die Auswirkung von Manipulationen der internen Struktur auf die Performanz vergleichen bzw. den Anteil partieller interner Strukturen an verschiedenen Aspekten der Performanz untersuchen. Hierzu zählen die Sensitivitätsanalyse, die Contribution Analysis sowie die Diskriminanzanalyse. Ein großer Vorteil gegenüber den zuvor vorgestellten Verfahren ist, dass in diesen Analysen die Bedeutung der projizierenden Gewichte zum Output-Layer und damit der unterschiedliche Anteil einzelner Units oder Gewichte an bestimmten Performanz-Aspekten ermittelt werden kann. Beispielhaft sei die Sensitivitätsanalyse genannt, die die Untersuchung komplexer Systeme ermöglicht. Dabei werden bei ansonsten gleichbleibenden Bedingungen einzelne Parameter variiert, um festzustellen, welche Veränderung zu einer signifikanten Änderung des Systemverhaltens führt. In Bezug auf die Analyse konnektionistischer Netzwerke heißt das, bedeutungsvolle Unterschiede im Repräsentationsraum durch die systematische Partitionierung oder Läsionierung (Entfernung einzelner Units oder Gewichte) des Netzwerks zu erfassen (vgl. Botvinick & Plaut, 2004; Allen & Seidenberg, 1999; Bullinaria, 1997; Christiansen & Chater, 1999b).

#### 3.1.4. Die Frage der Systematizität der Verarbeitung

SRNs verfügen über Mechanismen, mit denen grundlegende Erfordernisse geleistet und Einwände zurückgewiesen werden können, die den Spracherwerb betreffen. Eine weitergehende grundlegende Debatte, die Gegenstand vieler Studien wurde, betrifft die Systematizität, die durch SRNs abgebildet werden kann. Die Systematizität der Performanz betrifft den Grad der Generalisierungsfähigkeit von gelernten auf neue Strukturen. Dies erfordert die Erfüllung des Kriteriums, dass ein Modell auf eine Weise so über syntaktische Strukturen generalisieren können muss, dass es die Struktur gleichermaßen mit neuen Konstituenten interpretieren kann, auch wenn diese Realisierung von Struktur und Konstituenten noch nie gesehen wurde. Dies bedeutet, dass die strukturelle Realisierung unabhängig von den erlebten Realisierungen interpretierbar sein muss. Das klassische Beispiel dazu kann am Satz *Jonathan liebt Simone* nachvollzogen werden: Ein kompetenter Sprecher des Deut-

schen, der diesen Satz verstehen kann, versteht auch zwingendermaßen den Satz *Simone liebt Jonathan*, selbst wenn er ihn noch nie gesehen hat.

Es wird allgemein angenommen, dass nicht nur die Sprache, sondern auch verschiedene weitere Gegenstandsbereiche eine Systematizität der Verarbeitung erfordern, und dass das kognitive System des Menschen einen hohen Grad an Systematizität leisten kann. Von mehreren Autoren wurde demgegenüber behauptet, dass konnektionistische Modelle – wenn überhaupt – nur eine beschränkte Systematizität aufweisen. Dies ist eine fundamentale Kritik, da sie den Status des Konnektionismus als Rahmentheorie menschlicher Kognition in Frage stellt. Sehr prominent wurde sie von Fodor und Pylyshyn (1988) formuliert: In ihrer Analyse argumentieren sie, dass nur symbolische Systeme Repräsentationszustände mit kompositionaler syntaktischer und semantischer Struktur aufweisen (vgl. Kap. 2.2.2). Da konnektionistische Systeme keine kompositionalen Repräsentationen entwickeln, können sie keine Systematizität zeigen – es sei denn, sie implementieren ein klassisches Symbolsystem.

Diese Kritik wurde in späteren Studien konkretisiert, indem verschiedene Stufen der Systematizität unterschieden wurden, die operationalisiert und empirisch geprüft werden können. So unterscheidet Hadley (1994) zwischen drei Stufen der Systematizität, die er anhand des Unterschieds zwischen Trainings- und Test-Set festlegt, über die ein Netzwerk erfolgreich generalisieren muss: (1) Schwache Systematizität gilt, wenn das Test-Set nur neue Sätze mit syntaktischen Strukturen enthält, die es im Trainings-Set gesehen hat, und nur Wörter auf Positionen enthält, auf denen diese im Trainings-Set erschienen sind. (2) Quasi-Systematizität gilt, wenn das Test-Set darüber hinaus nur Sätze mit komplexer syntaktischer Struktur (eingebettete Sätze) enthält, die im Trainings-Set zumindest in isomorpher Form in einfachen Sätzen gesehen wurden. (3) Starke Systematizität gilt, wenn das Test-Set in den neuen einfachen und komplexen Sätzen Wörter auf syntaktischen Positionen enthält, auf denen sie im Trainings-Set nie gesehen wurden. In einer Erweiterung und Schärfung dieser Systematik schlagen Niklasson und van Gelder (1994) eine Einteilung mit fünf Levels der erforderlichen Systematizität vor: (0) Keine Neuheit, jeder Testsatz erscheint im Trainings-Set, keine Generalisierung erforderlich. (1) Neue Formel, aber die atomaren Komponenten erscheinen nur auf den Positionen, auf denen sie auch im Trainings-Set standen. (2) Neue Positionen, mindestens eine atomare Komponente erscheint in einer syntaktischen Position, an der sie noch nie zuvor erschienen ist. (3) Neue Konstituenten, mindestens eine atomare Komponente wurde noch überhaupt nicht gesehen. (4) Neue Komplexität, Testsätze enthalten einen höheren Komplexitätsgrad, zum Beispiel eine größere Einbettungstiefe. (5) Neue Konstituenten in neuer Komplexität, Kombination einer nie gesehenen Konstituente auf einer Position innerhalb eines neuen Komplexitätsgrads. Unabhängig von der Frage der Systematizität in konnektionistischen Systemen halten Niklasson und van Gelder (1994) dabei allerdings fest, dass das erforderliche Level der Systematizität zur Modellierung menschlicher Performanz erst noch empirisch im Detail zu klären bleibt und zudem auch bislang unklar ist, ob klassische Architekturen dieses Level der Systematizität erfüllen können.

#### 3.1.4.1. Systematizität von SRNs

In verschiedenen Studien wurde die Generalisierungsleistung konnektionistischer Architekturen detailliert in Frage gestellt. (Hadley, 1994) untersucht in einem Review die Systematizität der Performanz mehrerer konnektionistischer Architekturen – darunter SRNs (Elman, 1990), das SG-Modell von St. John und McClelland (1990), RAAMs (Pollack, 1990), das auf diesen basierende Modell von (Chalmers, 1990) und die Tensor-Produkt-Konzeption von Smolensky (1990). Dabei kommt er zu dem Schluss, dass diese Modelle nicht die Stufe starker Systematizität erreichen, die er als Voraussetzung für die Modellierung menschlicher Kognition annimmt (vgl. auch Hadley et al., 2001). Speziell für SRNs wird dabei attestiert, dass ihre Systematizität nur die zweite Stufe erreicht und dies nur innerhalb sehr limitierter Korpora mit sehr kleinem Vokabular gegeben und sehr fragil ist. Van der Velde, Van der Voort van der Kleij und De Kamps (2004) bestreitet sogar die Kapazität von SRNs zur Erfüllung schwacher kombinatorischer Systematizität und demonstriert entsprechende Einschränkungen mit einer Simulationsserie. R. Frank und Mathis (2007) kritisieren die Systematizität des Erwerbs von Sprache in SRN-Modellen. Dabei problematisieren sie den angeblichen Einfluss irrelevanter nicht-struktureller Faktoren auf die strukturelle Kategorisierung und vor allem die Unfähigkeit von SRNs, irrelevante lineare Generalisierungen zu ignorieren.

**Systematizität über syntaktische Klassen** In mehreren Studien wurde versucht, die Systematizität von SRNs mit Vorhersage-Aufgabe zu belegen (u.a. Elman, 1991, 1993; Christiansen & Chater, 1994; Brakel & Frank, 2009). Dabei wurden systematisch Konstituenten für verschiedene Satzpositionen vorenthalten und anschließend die Generalisierung für diese getestet. Die Befunde, die hierzu gezeigt werden, sind allerdings wenig eindeutig. So können etwa die SRNs in der Studie von Christiansen und Chater (1994) entsprechend vorenthaltenen Nomina nicht auf die Position des Genitiv-Objekts generalisieren, eingeschränkt aber auf die Position des Akkusativ-Objekts. Ein ähnliches Vorgehen erläutert Elman (1998): In diesem wurde für ein einzelnes Nomen gezeigt, dass es nach einer anfänglich geringeren Aktivierung nach ausreichendem Training auch auf der vorenthaltenen Position in gleichem Maß aktiviert wird, wie die anderen Nomen seiner Klasse. Bei dieser Form der Generalisierung stellt sich auch immer die Frage, ob eine Lücke nur zufällig oder intendiert ist. Denn es ist auch ein typisches Phänomen natürlicher Sprachen, dass es systematische Ausnahmen gibt, nach denen ein Wort – obwohl es zu einer Klasse gehört, die auf einer bestimmten syntaktischen Position erscheinen kann – auf einer kategorial passenden syntaktischen Position ungrammatisch ist. Wie können SRNs ohne spezifisches vorstrukturiertes Wissen zwischen zufälligen und intendierten Lücken unterscheiden? Elman zeigt, dass das Netzwerk bei weiterem Training eine Schwelle überschreitet und die vorenthaltenen syntaktische Position für das Nomen als Ausnahme erkennt. Der Unterschied wird also als Effekt der Erfahrung erklärt. Ob dies eine befriedigende Erklärung sein kann, bleibt zu diskutieren, soll aber an dieser Stelle nicht weiter verfolgt werden.

Erstaunlich erfolgreich scheint im Vergleich zu den frühen Ansätzen die Studie von Brakel und Frank (2009). In ihren Simulationen erschien die Klasse femininer Nomina nur als Satz-Subjekt und die Klasse maskuliner Nomina nur als Satz-Objekt. Als überbrückende Klasse, die auf beiden Positionen erscheinen kann, wurden Tier-Nomen in

die Grammatik eingefügt. Dennoch zeigten die SRNs für die Testsätze – komplexe Sätze mit eingebetteten Subjekt- und Objekt-Relativsätzen, in denen maskuline und feminine Nomina die Positionen belegen, auf denen sie während des Trainings nicht gesehen wurden – eine erstaunlich gute Performanz. Leider wurde nur die globale Performanz auf Grundlage des Cosinus zwischen der Wahrscheinlichkeit laut PCFG und dem Vektor der Output-Aktivierung dargestellt (Probleme globaler Fehlermaße und speziell des Cosinus besprechen wir in Kapitel 3.2.1.1). Eine starke Systematizität schließen die Autoren in erster Linie aus dem Vergleich mit Hidden Markov-Modellen, die selbst von den schlechtesten SRNs der Simulationsserie auf den kritischen Positionen übertroffen wurden. Das Fehlen von Detailanalysen der Aktivierungen und das Fehlen von Vergleichen mit Sätzen mit der Positionsverteilung des Trainings-Sets machen eine endgültige Beurteilung des Erfolgs dieser Simulationen unmöglich. Sehr viel problematischer ist an dieser Studie jedoch die Gegebenheit, dass die Grammatik keine Kongruenz zwischen Elementen der Sprache (z.B. Subjekt-Prädikat-Kongruenz) erfordert. Dies legt nahe, dass der Erfolg der Simulation nur daher rührt, dass die SRNs keine Fernabhängigkeiten abbilden mussten und die gezeigten Generalisierungen deshalb auf kleinsten Übergangsbereichen beruhen dürften.<sup>17</sup>

Insgesamt zeigen die SRNs in allen bisherigen Studien eine hohe Fragilität der gezeigten Systematizität. Auch Elman (1998) stellt fest, dass es kritische Effekte der Größe und der Struktur der Trainingskorpora sowie des Trainingsumfangs gibt, die noch gründlicher Untersuchung bedürfen. Hadley et al. (2001) bemängeln weitergehende Schwächen der verschiedenen Studien (v.a. bei Elman, 1998; Christiansen & Chater, 1994): Unter anderem vermuten sie, dass deren Befunde daher rühren, dass auf den vorenthaltenen Positionen ausschließlich Nomen erscheinen können. In eigenen Simulationen mit schwach nicht-deterministischen Grammatiken zeigen sie, dass die Vorhersage für vorenthalte Elemente nicht erfolgreich ist, wenn nach dem gegebenen syntaktischen Kontext außer Nomen auch andere Wortklassen folgen können. Hadley geht davon aus, dass die gefundenen Anzeichen von Systematizität auf den lexikalisch und grammatisch stark beschränkten Sprachkorpora und der geringen Anzahl vorenthaltener Elemente bzw. Positionsbesetzungen beruhen, und insgesamt kaum Robustheit besitzen.

Mit einer etwas realistischeren Perspektive auf die Möglichkeiten von SRNs untersucht S. L. Frank (2006b) die Systematizität der Satzverarbeitung mit rekurrenten konnektionistischen Netzwerken. Er kommt zu dem Ergebnis, dass SRNs zumindest eine schwache kombinatorische Systematizität aufweisen. Dabei haben sowohl die Erweiterung des Lexikonumfangs der zu lernenden Sprache als auch die Vergrößerung des Hidden-Layers einen positiven Effekt auf die Generalisierungsleistung – innerhalb einer nach oben und unten beschränkten Spanne. Frank erklärt dies durch Kapazitätsbeschränkungen in zu kleinen und Overfitting (Auswendiglernen von Abfolgen statt Generalisierung) in zu großen Netzwerken. Um dieser Grenze nach oben zu entkommen, schlägt er eine starke Vergrößerung des rekurrenten Layers mit einer geringen Verknüpfungsdichte mit nicht trainierbaren Gewichten vor. Er beschreibt dies als ein System mit sehr großem Kurzzeitgedächtnis und sehr kleinem Langzeitgedächtnis. Diese Konfiguration ist zum Beispiel bei

---

<sup>17</sup>Nichtsdestotrotz weist die Input-Gewichtsmatrix eine anhand hierarchischer Clusteranalyse abgebildete Kategorisierungsstruktur auf, die eine grundlegende Ableitung der Wortklassen und Zuordnung der Tokens zeigt. Die Autoren weisen hierbei darauf, dass die Nomen trotz unterschiedlichem Genus zusammen in ein Cluster fallen und führen die korrekte, von der Trainingsposition unabhängige, Aktivierung darauf zurück.

den sogenannten Echo State Networks (ESNs) gegeben, mit denen Frank die Annahme überprüft, ob in einem solchen System das Overfitting-Problem überwunden werden kann. Tatsächlich stellt er einen monoton verbessernden positiven Effekt der Vergrößerung des rekurrenten Layers in solchen Netzwerken fest.<sup>18</sup> Neben ESNs ist das RecSOMsard als alternative rekurrente Architektur zu nennen: Farkaš und Crocker (2006) entwickeln damit eine selbstorganisierte Alternative zu SRNs, in der die Vorhersage-Aufgabe mit unüberwachten Trainingsalgorithmen kombiniert wird. Während die Architektur oberflächlich die gleiche Struktur aufweist wie SRNs, werden für das Hidden- und das Kontext-Layer selbstorganisierte Karten (SOM, Kohonen, 1990) verwendet. In einer Untersuchung der syntaktischen Systematizität präsentieren Farkaš und Crocker (2008) positive Befunde, und weisen damit den Kreis rekurrenter Architekturen aus, die ein in dieser Hinsicht stabiles Verhalten aufweisen.

**Systematizität bei der Variablenbindung** Einen etwas anderen Aspekt der Systematizität problematisiert G. F. Marcus (1998). Dabei bezieht er sich speziell auf die Beschränkungen eliminativer (im Sinne von nicht symbolsystem-implementierender) konnektionistischer Modelle. Er attestiert diesen Modellen eine Unfähigkeit zur Abbildung der für die menschliche Kognition grundlegende Funktion der ad hoc-Variablenbindung als kontextunabhängige Generalisierung unbekannter Objekte/Sachverhalte im Sinne einer nicht gelernten spontanen Zuordnung an einen Stellvertreter. Dies macht er an der Identitätsfunktion fest, die eine ad hoc-Zuweisung unbekannter Elemente leistet (z.B. *Ein X ist ein X*). Speziell zu SRNs konstatiert er: „The simple recurrent network, however, is incompatible with generalizing a particular class of universals, *universally quantified one-to-one mappings* [Hervorhebung im Original].“ (G. F. Marcus, 1998, Seite 267). Die fehlende Fähigkeit, diese Form der Generalisierung zu leisten, führt (G. F. Marcus, 1998) auf intrinsische Eigenschaften der Netzwerke und ihrer Lernmechanismen zurück, die dazu führen, dass diese Netzwerke nicht außerhalb des Trainingsereignisraums generalisieren können: Die Eigenschaft der Trainings-Unabhängigkeit beruht auf den lokalistischen Fehleranpassungsmechanismen der Gewichte und führt dazu, dass Gewichte zwischen Aktivierungsereignissen nicht angepasst werden, wenn ihre Assoziation nicht trainiert wird.<sup>19</sup> Es geht aber nicht nur um die absolute Neuheit einer Konstituente, sondern auch um die Neuheit der Konstituente in einer bestimmten Funktion (G. F. Marcus, 1998, Seite 269):

What matters is the training space *with respect to some particular function* [Hervorhebung im Original]. Any item that is outside the training space with respect to that function – even if it is within the training space of some other function – will not be generalized properly.

---

<sup>18</sup>Allerdings weisen ESNs andere Schwächen auf, die den Autor dazu bewogen haben, von dieser Architektur als Grundlage für die Modellierung von Sprachverarbeitung wieder Abstand zu nehmen (S. L. Frank, 2009, persönliche Kommunikation).

<sup>19</sup>Dies führt (G. F. Marcus, 1998) auf die spezielleren Eigenschaften der Input-Unabhängigkeit und der Output-Unabhängigkeit im Backpropagation-Algorithmus zurück. So werden zum einen die Gewichte einer nicht aktivierten Input-Unit nicht bei der fehlerabhängigen Anpassung der Gewichte berücksichtigt. Zum anderen werden die Gewichte der verschiedenen Output-Units gänzlich unabhängig voneinander angepasst. Das Problem liegt also nicht an einem Fehler der in den internen Aktivierungsmustern abgebildeten Klassifizierungen, sondern an der nicht stattfindenden Anpassung der für die betreffende Unit relevanten Gewichte.

Das heißt, selbst wenn die Konstituente im Trainingsereignisraum erschienen ist und in ansonsten jeglicher Hinsicht – außer einer spezifischen Funktion – kategorial korrekt zugewiesen wurde, so liegt aber ihre Verwendung in dieser ausgelassenen Funktion außerhalb des Trainingsereignisraums und kann von konnektionistischen Netzwerken nicht korrekt erworben werden. Um Funktionen und Zuweisungen zu extrapolieren, müssen Mechanismen zur Symbolmanipulation implementiert werden.

Elman (1998) weist die Kritik an der Leistungsfähigkeit von SRNs durch G. F. Marcus (1998) entschieden zurück. Er demonstriert, dass SRNs sehr wohl auch auf neue Stimuli generalisieren können. Die Kritik, dass sie aber nicht außerhalb des Trainingsereignisraums generalisieren können, kommentiert er hingegen als Fakt, der in der menschlichen Kognition gleichfalls gegeben ist (Elman, 1998, S. 16):

A network which is not designed to perceive along a given input dimension, or which is systematically deprived of experience in that dimension, will fare no better than a human who has not evolved to see in the infrared, or a cat who is reared in an environment containing no horizontal stripes.

Wir stimmen zu, dass eine Wahrnehmungsdimension, die systematisch vorenthalten wird, auch vom Menschen nicht verarbeitet werden kann. Allerdings scheint es, als wolle Elman Marcus hier absichtlich missverstehen, indem er seine Diskussion auf den absoluten Fall nie trainierter Units reduziert. Auf die spezifischere Aussage, dass dies auch dann der Fall ist, wenn eine Unit nur in einer bestimmten Funktion vorenthalten wird – zum Beispiel eine Unit, die ein Nomen repräsentiert, das nur im Nominativ und im Akkusativ erscheint, nie aber im Genitiv – geht Elman überhaupt nicht ein. Dies ist in der Sprachverarbeitung aber die Regel, keine Ausnahme. R. Frank et al. (2005) gehen näher auf dieses Problem der Identitätsfunktion in SRNs ein und untersuchen die Fähigkeit dieser Architektur zur Auflösung anaphorischer Referenzen. Für die Verarbeitung von Reflexivpronomen ist diese Funktion direkt relevant. Ihre Ergebnisse unterstützen die Annahmen von (G. F. Marcus, 1998) insofern, dass ihre Netzwerke diese Aufgabe bei entsprechend vorenthaltenem Referenten nicht leisten konnten. Allerdings kann eine misslungene Modellierung sicherlich nicht als Gegenbeweis akzeptiert werden. Wir gehen in Kapitel 6 noch näher auf diese Studie ein.

### 3.1.4.2. Systematizität anderer konnektionistischer Architekturen

Natürlich wurden nicht nur SRNs in ihren Eigenschaften zur Abbildung von Systematizität untersucht. Die Konzentration auf die Eigenschaften von SRNs hatte sicherlich damit etwas zu tun, dass sie die meistverwendete Architektur für die Verarbeitung sequenzieller Strukturen ist und für die konnektionistische Modellierung von Sprachverarbeitung von Beginn an eine herausragende Bedeutung entwickelt hat. Andere Architekturen haben dabei mitunter ein wesentlich höheres Level der Systematizität erreicht. Deren Leistungen beruhen aber in der Regel auf einer Implementierung von Aspekten klassischer Symbolsysteme, die durch die Trennung von Wissen und Verarbeitung, die Verwendung explizit enkodierter Variablenbindung oder expliziter semantischer Repräsentationen zustandekommen, die nach dem Vorbild symbolischer Modelle vorstrukturiert sind. Damit entsprechen diese Ansätze der Annahme von Fodor und Pylyshyn (1988), dass konnektionistische Modelle nur dann systematisches Verhalten aufweisen können, wenn sie ein klassisches Symbolsystem

implementieren. Wir besprechen kurz einige Ansätze in dieser Tradition, werden dabei aber nicht näher auf Details eingehen.

Niklasson und van Gelder (1994) beschreiben ein Modell, das RAAMs (Pollack, 1990, vgl. Kap. 3.1.1) und sogenannte Transformation Networks<sup>20</sup> kombiniert. Ein deutlicher Nachteil ist, dass dieses Modell die Vorgabe explizit kodierter syntaktischer Klassen erfordert. Auf Seiten der Systematizität beanspruchen die Autoren jedoch, dass das Modell zumindest das dritte Level ihrer Systematisierung erreicht (Verarbeitung gänzlich neuer Konstituenten, vgl. oben). Diese Leistung wird durch die in konnektionistischen Modellen unübliche Trennung von Wissen und Verarbeitung erreicht, die das Modell eher in die Klasse symbolischer Modelle rückt bzw. der Realisierung einer konnektionistischen Implementierung symbolischer Verarbeitungsmechanismen entspricht. Ebenfalls eine Lösung im Sinne des implementierenden Konnektionismus entwickeln Hummel und Holyoak (2001); Hummel et al. (2004) mit ihrem LISA-Modell (LISA: Learning and Inference with Schemas and Analogies). In diesem werden explizite Variablen und Beziehungen zwischen Variablen ausgewiesen, wodurch das Modell problemlos dazu in der Lage ist, Ad-hoc-Variablenbindung und somit auch die Identitätsfunktion abzubilden.

Auch Dominey und Ramus (2000) vermuten, dass die Fähigkeiten zur Extraktion statistischer Regularitäten und zur Variablenbindung auf verschiedenen kognitiven Mechanismen beruhen. Ihren Modellierungen liegt die Beobachtung zugrunde, dass bereits Säuglinge sensitiv für serielle, temporale und abstrakte sprachliche Strukturen sind. Im Zentrum steht die Frage, welche Mechanismen der schnellen Lernfähigkeit für diese Strukturen zugrunde liegen und welche Rolle die vorhergehende Erfahrung mit Umweltreizen spielt. Die Verarbeitung sequenzieller und rhythmischer Strukturen (Phonologie) wird in ihren Simulationen durch Temporal Recurrent Networks (TRNs, vgl. Kap. 3.1.1) geleistet. Das Lernen abstrakter Strukturen setzt allerdings eine Erweiterung um ein Abstract Recurrent Network voraus, einer Implementierung eines Kurzzeitspeichers mit Rekognitionsfunktion, die die Verarbeitung von Wiederholungsstrukturen und damit die Erfassung der Identitätsfunktion leistet.

Dem Ansatz von Hadley et al. (2001) liegt die Annahme zugrunde, dass die Systematizität der syntaktischen Ebene aus der Vorhersagbarkeit auf semantischer Ebene entsteht: Die Grundlage dieser menschlichen Leistungsfähigkeit beruht auf der Inferenzfähigkeit aus tiefgreifendem semantischem Wissen. Für ihre Simulationen verwenden sie ein kompetitives hebbisches Netzwerk, dessen Architektur zwei Kontextschichten beinhaltet. Eine Ähnlichkeit zu SRNs wird von den Autoren nicht bestritten. Der wesentliche Unterschied zu SRNs mit Wortvorhersage-Aufgabe besteht darin, dass das Netzwerk im Kontext der semantischen Merkmale des gegenwärtigen Wortes die semantischen Merkmale des Folgewortes vorhersagen soll. Dem Netzwerk werden zur Verarbeitung explizit vorstrukturierte semantische Repräsentationen gegeben. Dieses Netzwerk erfüllt beim Testen auf systematisch vorenthaltene Positionen das Kriterium starker Systematizität.

Zu Recht kritisieren S. L. Frank et al. (2009), dass das Modell von Hadley et al. (2001) seine Systematizität den semantischen Repräsentationen verdankt, die es eher zu einer Instantiierung eines klassisch-symbolischen Modells machen. Auch in ihrem eigenen Modell

---

<sup>20</sup>Sofern dies aus den Angaben von Niklasson und van Gelder (1994) erkennbar ist, handelt es sich beim Transformation Network um ein einfaches einschichtiges FFN.

(vgl. u.a. S. L. Frank, Koppen, Noordman & Vonk, 2003b, 2007, 2008) spielt die Inferenz aus semantischem Wissen eine bedeutende Rolle bei der Erzeugung systematischen Verhaltens. Allerdings legen sie großen Wert darauf, dass sie in ihrem Modell keine symbolischen Strukturen im Sinne explizit kategorial vorstrukturierten semantischen Wissens implementieren. Entgegen der Behauptung von Fodor und Pylyshyn (1988) nehmen Frank und Kollegen an, dass konnektionistische Modelle sehr wohl die Eigenschaft der Systematizität besitzen, diese allerdings nicht auf den internen Eigenschaften beruht, sondern robust aus der Beziehung zur systematischen Struktur in der Welt entsteht, die durch die Modelle erfasst und verarbeitet wird. Das Modell von S. L. Frank et al. (2009) entwickelt starkes systematisches Verhalten, das in verschiedenen, gut ausgearbeiteten Kategorien getestet wurde. Wir werden uns in den späteren Kapiteln 4 und 9 noch ausführlicher mit diesem unseres Erachtens bislang überzeugendsten Ansatz zur Untersuchung konnektionistischer Systematizität befassen.

#### 3.1.4.3. Konnektionistische Systematizität

Fassen wir zusammen: Zur Beantwortung der Systematizitätsfrage wurden in der Forschung verschiedene Strategien verfolgt. Zum einen wurde versucht zu zeigen, dass SRNs die erforderliche Systematizität besitzen. Zum anderen wurde versucht zu überprüfen, welche Systematizität überhaupt erforderlich ist, ob die Annahmen über die Systematizität des menschlichen Vorbilds realistisch sind, und ob die daraus abgeleiteten Anforderungen an SRNs nicht schlicht zu hoch sind. Speziell für SRNs mit Wortvorhersage-Aufgabe gibt es zwar erste Ansätze zur Demonstration systematischen Verhaltens, allerdings weisen die Befunde darauf hin, dass das Verhalten nicht sehr robust ist. Die Lösung der Identitätsfunktion scheint überhaupt nicht zu gelingen. Bessere Ergebnisse erzielen Modelle, die Erweiterungen in ihre Architektur aufnehmen, die Mechanismen symbolischer Systeme oder kombinatorisch strukturierte Repräsentationen implementieren. Aus dieser relativierenden Perspektive des implementierenden Konnektionismus gehen allerdings andere Aspekte verloren, die den Konnektionismus als neue alternative Perspektive erst interessant gemacht haben – beispielsweise die Aufhebung der klaren Distinktion zwischen Wissensstrukturen und Verarbeitungsprozessen, die Ablehnung der Annahme expliziter diskreter Wissensstrukturen und algorithmisch-syntaktischer Verarbeitungsmechanismen. Um so wichtiger ist es, im Sinne von (S. L. Frank et al., 2009) Modelle zu entwickeln, die robuste Systematizität in konnektionistischen Netzwerken belegen, die explizit kein klassisch-symbolisches Modell implementieren. Das SRN mit klassischer Wortvorhersage-Aufgabe scheint jedoch nach den gegenwärtigen Befunden keine geeignete Methode zur Modellierung und Erklärung starker Systematizität zu sein – sofern man Systematizität von Sprache und Kognition als tatsächliche empirische Tatsache akzeptieren möchte, was zum Beispiel von Johnson (2004) bezweifelt und kritisch diskutiert wird.<sup>21</sup>

---

<sup>21</sup>Johnson (2004) demonstriert in einer kritischen Analyse, dass die Existenz von Systematizität in der verbreitet angenommenen Eindeutigkeit nicht feststeht bzw. nicht in der simplen Form gegeben ist. Im Falle der Systematizität natürlicher Sprachen besteht das Problem, dass die Demonstration ihrer Existenz auf Grundlage theoretisch angenommener Konstituentenklassen falsch ist, da in jeder Klasse Ausnahmen existieren, für die die systematische Verwendung innerhalb von Konstruktionen ungrammatisch wäre. Um Systematizität einer Sprache tatsächlich empirisch erfassen zu können, werden viel feinkörnigere natürliche Klassen gefordert, die aber wiederum das Problem der Klassenüberlappung (gegenüber einer

Dennoch eignen sich SRNs mit Wortvorhersage-Aufgabe für eine Vielzahl anderer beobachteter Eigenschaften menschlicher Sprachverarbeitung. Und in der Erklärungskraft, die das Modell hier besitzt, scheint es eine gute Grundlage für weitergehende und umfassendere Modelle zu sein, die einen größeren Ausschnitt empirischer Realität abbilden. Die Berücksichtigung semantischer Inferenzen als Grundlage für die Systematizität syntaktischer Verarbeitung, wie sie von Hadley et al. (2001) oder S. L. Frank et al. (2009) vorgeschlagen wird, scheint ein guter Ansatz für eine entsprechende Erweiterung zu sein. Innerhalb des empirischen Ausschnitts, der mit SRNs untersucht werden kann, ist die Grundarchitektur aber nach wie vor ein überzeugendes Modell, dessen Erklärungsmächtigkeit unseres Erachtens nicht durch die Beschränkungen in der Systematizitätsfrage beeinträchtigt wird.

### 3.2. SRNs als psycholinguistisches Modell

In den vorhergehenden Abschnitten wurde die grundlegende Leistungsfähigkeit von SRNs skizziert – die prinzipielle Erwerbbarkeit komplexer syntaktischer Strukturen über rein induktive assoziative Mechanismen. In den folgenden Abschnitten betrachten wir die Performanz von SRNs – ihre Verwendbarkeit als deskriptive und prädiktive Modelle menschlicher Performanz bei der natürlichsprachlichen Satzverarbeitung. Welche Vorhersagen werden mit ihnen für Verarbeitungsfehler, Lesezeiten und Aspekte von Komplexität und Ambiguität getroffen? Wir diskutieren eine Reihe von Simulationsstudien, die Satzverarbeitung unter Verwendung von SRNs mit der Vorhersage-Aufgabe modellieren. In diesen Studien wird untersucht, welche neuen Perspektiven und Einsichten ein solches Multiple-Constraint-Satisfaction-Modell mit dynamischem Gedächtnis auf die Online-Satzverarbeitung eröffnet, die hier als syntaktisch-lexikalische Projektion verstanden werden kann. Dabei wird zu klären sein, wie die relevanten Aspekte menschlicher Satzverarbeitungsperformanz in der Modellierung operationalisiert, erklärt und quantifiziert werden. Unter den Eigenschaften von SRNs, die als Erklärungsprinzipien der menschlichen kognitiven Kapazität zur Sprachverarbeitung herangezogen werden, sind die folgenden hervorzuheben:

- Projektion bzw. Antizipation als eine grundlegende Aufgabe der Sprachverarbeitung
- Interpretation der mentalen Grammatik als System von erfahrungsbasierten Attraktoren in einem dynamischen System, in dem syntaktische Strukturen als Trajektorien durch den hochdimensionalen Zustandsraum beschrieben werden können
- Inputsensitivität und Kontextsensitivität, die zu Effekten und Interaktionen von Frequenz und Regularität der Inputstruktur – Items, Item-Kollokationen, syntaktische Strukturen und Item-Struktur-Kollokationen – führen
- Sensitivität für multiple Constraints, die die Extraktion und Nutzung statistischer Cues bei Erwerb und Verarbeitung von Regularitäten ermöglicht und eine massive Parallelverarbeitung und Interaktivität unterschiedlicher Informationsquellen vorgibt

---

distinkten Partition) haben. Bevor Modelle näher nach ihrer Systematizität bei der Abbildung einer spezifischen Sprache evaluiert werden können, müssten eigentlich zuerst die empirischen Fakten über deren natürliche Klassen bestimmt werden.

- Interferenzen durch beschränkte Kapazität und Aktivierungsüberlagerung, die in Effekten der Sensitivität für lokale Strukturen resultieren und Einschränkungen bei der Verarbeitung komplexer Strukturen und Fernabhängigkeiten mit sich bringen

Diese Eigenschaften werden in vielen aktuellen empirischen Studien in Bezug auf Phänomene der Frequenz und Erfahrungsabhängigkeit, der Interferenz, der Lokalität und Anti-Lokalität sowie der lokalen Kohärenz thematisiert.

### 3.2.1. Beschränkte Kapazität, Interferenz und Antizipation

Wir haben bereits den Ansatz von Elman (1991) zur Untersuchung der Verarbeitungsfähigkeit von SRNs für komplexe Sätze mit eingebetteten Nebensätzen angesprochen. Doch welche Verarbeitungseigenschaften weisen SRNs für komplexe Satzstrukturen auf? Welche Verarbeitungsunterschiede zeigen sie für verschiedene komplexe Strukturen? Und welche Besonderheiten finden sich bei der empirischen Untersuchung der Online-Verarbeitung komplexer Satzstrukturen, die das konnektionistische Modell in irgendeiner vergleichbaren Form nachbilden sollte?

**Interferenzen offener Anbindungen** Diese Fragen betreffend wurden zumeist komplexe syntaktische Einbettungen untersucht. Beispiel 3 führt Sätze mit verschiedenen Typen von Nebensatz-Einbettung auf: Satz 3a zeigt einen rechtsverzweigenden Relativsatz, der attribuerend an das Objekt des Matrixsatzes angehängt wird. Satz 3b zeigt einen zentral eingebetteten Relativsatz, der attribuerend an das Matrixsatz-Subjekt angehängt wird, dasselbe ist bei Satz 3c der Fall. Die beiden Sätze unterscheiden sich darin, welche Satzfunktion das attribuierte Element im Relativsatz erhält: Im Subjekt-extrahierten Relativsatz (auch kurz: Subjekt-Relativsatz, Bsp. 3b) wird das durch das Relativpronomen aufgegriffene Matrixsatz-Subjekt zum Subjekt des Relativsatzes. Im Objekt-extrahierten Relativsatz (auch kurz: Objekt-Relativsatz, Bsp. 3c) wird es zum Objekt des Relativsatzes. Satz 3d zeigt eine multiple Einbettung von Relativsätzen, d.h. zentral eingebettete Relativsätze, die wiederum in zentral eingebettete Relativsätze eingebettet sind. Der Satz weist eine Einbettungstiefe von drei Ebenen auf. Die letzten beiden Beispiele demonstrieren schematisch einen typologischen Unterschied bei der Relativsatzeinbettung: Satz 3e entspricht der geschachtelten Zuordnung von Subjekt und Prädikat und damit auch der geschachtelten Auflösung der Subjekt-Prädikats-Kongruenz zwischen den dependenten Elementen des Satzes bei der Zentraleinbettung (im Sinne eines last-in-first-out Speicher-Algorithmus). Dies entspricht der Dependenzordnung, wie sie etwa im Deutschen oder im Englischen gegeben ist. Satz 3f stellt dem die gekreuzte Zuordnung und Auflösung der dependenten Elemente gegenüber, in der die Elemente im Sinne eines first-in-first-out Speicher-Algorithmus angeordnet werden (vgl. auch Christiansen & Chater, 1999b). Dieses wesentlich seltenere Muster findet sich zum Beispiel im Schweizer-Deutsch und im Niederländischen. Eine frühe empirische Studie zur Verarbeitung verschiedener Strukturen, Anordnungen und Tiefen der Nebensatz-Einbettung mit Verständlichkeitsratings und Verstehensaufgaben wurde von Bach, Brown und Marslen-Wilson (1986) durchgeführt.

- (3) a. Der Bankier lobt den Polizisten, der den Hampelmann verhaftet.

- b. Der Polizist, der den Hampelmann verhaftet, ist gereizt.
- c. Der Hampelmann, den der Polizist verhaftet, ist auch gereizt.
- d. Als die Katze, die die Maus, die der Hund, den der Elefant beschnüffelt, beisst, ignoriert, auf die Ablage springt, ist das schöne Sonntagsgeschirr dahin.
- e.  $A_N B_N B_V A_V$
- f.  $A_N B_N A_V B_V$

Wir haben im vorausgegangenen Unterkapitel dargestellt, dass SRNs komplexe syntaktische Strukturen verarbeiten und komplexe Abhängigkeiten über diese Strukturen hinweg auflösen können (z.B. Elman, 1991). Ein erster Ansatz, um die dabei gezeigte Performanz zu analysieren und hinsichtlich ihrer Relevanz für empirische psycholinguistische Befunde zu untersuchen, findet sich bei Weckerly und Elman (1992). Die Simulationen sind auf die Frage der relativen Komplexität bzw. Verarbeitungspräferenz verschiedener Nebensatzstrukturen ausgerichtet. Speziell geht es um den Unterschied zwischen zentraleingebetteten und rechtsverzweigenden Relativsätzen (Simulation 1) sowie um den Einfluß von semantischem Bias auf die Vorhersagequalität (Simulation 2). Die Performanz wurde anhand des Cosinus (vgl. folgenden Exkurs, Kap. 3.2.1.1) zwischen den jeweiligen Aktivierungsvektoren und den Vektoren der absoluten Wortfolgewahrscheinlichkeiten über den Satzpositionen evaluiert. Dem liegt die Idee zugrunde, dass der Grad der Ab- und Zunahme des Antizipationsfehlers bzw. der Unterschied des Fehlers für kontrastierte Positionen als allgemeines Maß des Verstehens mit entsprechenden empirisch erfassten Performanzmaßen der menschlichen Satzverarbeitung verglichen werden kann. Eine präzisere Aktivierung im Modell sollte mit einem besseren Verstehen von Versuchsteilnehmern für die entsprechenden Sätze korrespondieren.

In Übereinstimmung zu behavioralen Daten zeigen die SRNs in der ersten Simulation eine bessere Performanz für Sätze mit doppelter Rechtsverzweigung mit Subjekt-Relativsätzen gegenüber doppelt zentraleingebetteten Objekt-Relativsätzen. Weckerly und Elman (1992) erklären das gefundene Muster für die Performanz und die relativen Präferenzen über die repräsentationalen Beschränkungen von SRNs. Diese Beschränkung wird im Unterschied zu klassischen Ansätzen nicht als Effekt separater nicht-sprachlicher Ressourcen verstanden (etwa als Beschränkung des allgemeinen Arbeitsgedächtnisses bei der Sprachverwendung), sondern als intrinsisches Merkmal des sprachverarbeitenden Systems. Die interne Repräsentation zu einem Zeitschritt enthält Informationen über das gegenwärtige Wort sowie über dessen vorausgegangenen syntaktischen Kontext und die hier eröffneten Abhängigkeiten. Sie indiziert damit die noch zu erwartenden Elemente, d.h. die offenen Argumentpositionen und Abhängigkeiten des Satzes, bevor dieser abgeschlossen werden kann. Verarbeitungsunterschiede resultieren aus den Unterschieden der Informationsmenge über die ausstehenden Argumente und Abhängigkeiten, die im Verlauf des Satzes aufrechterhalten werden müssen. Bei zentraleingebetteten Sätzen müssen im Zustandsvektor Informationen über Satzposition und gegenwärtige Einbettungstiefe sowie offene Satzsubjekte mitsamt deren grammatischen Merkmalen (z.B. Numerus für die Subjekt-Prädikat-Kongruenz) und semantischen Beschränkungen für mögliche Argumente aufrechterhalten werden. Bei rechtsverzweigenden Relativsätzen hingegen müssen nur die Informationen über das Satzsubjekt über eine sehr kurze Distanz aufrechterhalten werden. Die große Menge an aufrechtzuerhaltender Information bei Mehrfacheinbettung stellt

große Anforderungen an die repräsentationale Kapazität des Hidden-Layers, dessen Zustände an Präzision abnehmen, während die Wahrscheinlichkeit für Interferenzen zwischen Informationen erhöht wird.

Allerdings wurde an Details der Simulationen und Ergebnisse Kritik geäußert: So bemängelt Rohde (2002) zurecht, dass die Netzwerk-Performanz nicht zwingend einer Präferenz für Rechtsverzweigung gegenüber Zentraleinbettung von Sätzen in SRNs zugeschrieben werden kann. In den Sätzen der Studie sind sowohl im Trainings- als auch im Test-Set die Einbettungsposition der Relativsätze und der Extraktionstyp der Relativsätze (Subjekt-Relativsatz vs. Objekt-Relativsatz) konfundiert. Wie wir in den späteren Abschnitten noch diskutieren werden, wurde die dispräferierte Struktur von Objekt-Relativsätzen unter anderem auch auf ihre nicht-kanonische Wortstellung (im Englischen Objekt-Subjekt-Verb statt Subjekt-Verb-Objekt) zurückgeführt. Folglich muss berücksichtigt werden, dass die SRNs in der Studie eher eine Präferenz für Subjekt-Extraktion als für Rechtsverzweigung zeigen. Trotz dieser Kritik bleiben unseres Erachtens die allgemeinen Erklärungen zur Verarbeitung in SRNs relevant.

**Antizipation der Eigenschaften nachfolgender Wörter** Mit der zweiten Simulation untersuchen Weckerly und Elman (1992) den Einfluss semantischer Constraints. Die Ergebnisse werfen eine weitere Perspektive auf den Befund zur ersten Simulation: SRNs nutzen Informationen über semantische Einschränkungen für Co-Argumente, um das Spektrum möglicher Folgewörter einzuschränken und die Wortvorhersage dadurch zu präzisieren. In ihrem Beispiel verwenden die Autoren Beschränkungen für Verben in Abhängigkeit semantischer Merkmale des Nomens, bei denen zwischen belebt-menschlich und belebt-nicht-menschlich unterschieden wird. Die semantischen Informationen helfen dabei, auf der Position des Verbs für eine Menge aktiver Nomen das zum Verb gehörige Satzsubjekt zu bestimmen und durch die höhere Präzision der gegenwärtigen Vorhersage einen präziseren internen Zustand für die Vorhersage der Folgeposition zu erreichen. Semantische Constraints dienen also einer besseren Nutzung der beschränkten repräsentationalen Kapazität durch präzisere Informationsstrukturen und führen zu verbesserter Antizipation von erwarteten Folgewörtern.

Ähnlich sehen auch Konieczny und Döring (2003) eine zentrale Rolle der Antizipation für die Satzverarbeitung. Sie untersuchen mit einem Eyetracking-Experiment und einer Simulationsstudie mit SRNs die Rolle der Antizipation für die Verarbeitung satz-finaler Köpfe. Damit kontrastieren sie Theorien, die für die Verarbeitung satz-finaler finiter Verben in deutschen Nebensätzen einen Lokalitätseffekt vorhersagen: So wird zum Beispiel nach der Dependency Locality Theory (DLT, Gibson, 1998, 2000) angenommen, dass die Verarbeitung des Verbs mehr Ressourcen beansprucht und demnach umso komplexer ist, je höher die Anzahl der zuvor eingeführten Konstituenten ist, in deren Struktur das Verb integriert werden muss, und je weiter die Distanz zum Satzsubjekt ist, zu dem das Verb die Subjekt-Prädikat-Kongruenz erfüllen muss. Dem stellen Konieczny und Döring (2003) SRNs als probabilistisches vorhersagebasiertes Modell gegenüber: Je mehr Informationen durch zusätzlich gegebene Konstituenten zur Antizipation des ausstehenden Verbs gegeben sind, desto präziser sollte die Vorhersage des Verbs und seiner grammatischen Eigenschaften sein. Die Ergebnisse der Simulation zeigen eine deutlich bessere Vorhersage bzw. einen geringeren Vorhersagefehler für das Verb, je mehr Konstituenten ihm vorausgehen.

In einer darauf folgenden empirischen Studie wurden mit Blickbewegungsmessungen während des Lesens Lesezeiten aufgezeichnet. Die Sätze wurden so kontrastiert, dass vor dem satz-finalen finiten Verb in einem Relativsatz entweder Nominalphrasen im Dativ oder im Genitiv sowie zusätzlich eine das Nomen oder das folgende Verb modifizierende Präpositionalphrase vorausgehen. Nur die Dativ-NPs führen zusätzliche Information für die Antizipation des Verbs ein. Die Ergebnisse bestätigen die Befunde aus der Simulation: Trotz identischer Distanz zwischen Satzsubjekt und Verb fallen die Lesezeiten für das Verb nach vorausgehender Dativ-NP signifikant geringer aus. Die Präpositionalphrase hingegen hatte keinen Einfluß auf die Lesezeiten.

**Komplexität rekursiver Strukturen** Christiansen und Chater (1999b) stellen eine ausführliche Studie zur Verarbeitung rekursiver Strukturen in SRNs vor (vgl. auch Christiansen, 1994). Auf der Grundlage verschiedener Formen der Rekursion (in Anlehnung an Chomsky, 1957) erstellen sie unterschiedliche Sprachen, die jeweils eine Form von komplexer Konstruktion und rechtsverzweigender Konstruktion beinhalten. Im Zentrum stehen die beiden Sprachen mit zentraleingebetteten Relativsätzen (geschachtelte Dependenz zwischen Subjekten und Prädikaten, vgl. Bsp. 3e) sowie mit gekreuzt-abhängigen Relativsätzen (vgl. Bsp. 3f), kombiniert mit rechtsverzweigenden Relativsätzen (vgl. 3a). Ziel der Studie ist zum einen, die generelle Fähigkeit von SRNs zur Verarbeitung verschiedener rekursiver Strukturen zu belegen, zum anderen geht es darum, die Performanz der Netzwerke mit empirischen Daten zu vergleichen. Dafür werden vergleichende Evaluationen der Performanz der SRNs für die verschiedenen Formen der Rekursion und Ableitungen relativer Performanzwerte zum Vergleich mit empirisch erhobenen Daten aus Lesezeitmessungen und Grammatikalitätsratings durchgeführt. Hierfür schlagen (Christiansen & Chater, 1999b) das Performanzmaß Grammatical Prediction Error (GPE) vor. Der GPE berechnet sich über den Vergleich der probabilistischen Verteilung der Wortvorhersage im Output-Vektor mit der absoluten Wahrscheinlichkeit der Wortfolge laut der probabilistischen kontextfreien Grammatik (PCFG), aus der das Trainingskorpus erzeugt wurde (vgl. folgenden Exkurs, Kap. 3.2.1.1).

Die folgenden Befunde gehören zu den zentralen Ergebnissen aus den Simulationen: (a) Die komplexen Einbettungen führen in allen Fällen zu einem höheren Fehler als die rechtsverzweigende Rekursion. (b) Die Performanz der SRNs ist für die Sprache mit gekreuzten Abhängigkeiten besser als für die Sprache mit geschachtelten Abhängigkeiten. Dies äußert sich als Unterschied auf den kritischen Satzpositionen ab einer Einbettungstiefe von zwei Einbettungen, bei nur einer Einbettung hingegen findet sich kein Unterschied. (c) Mit steigender Anzahl der Einbettungen nimmt die Qualität der Performanz ab. Dies gilt unerwarteter Weise auch für rechtsverzweigende Einbettungen. Diese Befunde wurden mit den Ergebnissen aus verschiedenen psycholinguistischen Studien verglichen. Dazu wurde der Unterschied des GPEs für verschiedene Konstruktionen mit Unterschieden aus Grammatikalitätsratings, Verständnisfragen sowie Lesezeiten verglichen.

Der Befund zum relativen Vorteil von gekreuzten gegenüber geschachtelten Abhängigkeiten und zum Einsatz dieses Unterschieds erst ab der zweiten Einbettungstiefe deckt sich mit den Befunden von Bach et al. (1986). Diese vergleichen in ihrer Studie Ratings und Verständnisperformanz für gekreuzte Abhängigkeiten im Niederländischen und geschachtelte Abhängigkeiten im Deutschen. Auch der Einfluss der Einbettungstiefe auf die Verarbei-

tung von rechtsverzweigenden Sätzen wurde in dieser Studie untersucht – und auch hier findet sich eine deutliche Übereinstimmung zu den Daten aus der SRN-Simulation. Dieses Ergebnis steht im Kontrast zu Vorhersagen aus traditionellen symbolischen Modellen, die eine Kompetenz mit potenziell unendlicher Rekursivität annehmen, denn in diesen würde man für beliebige Tiefen der Rechtsverzweigung keine Performanzunterschiede erwarten. Ein weiteres interessantes empirisches Ergebnis der psycholinguistischen Forschung zeigt für das Englische, dass ungrammatische Sätze mit doppelter Zentraleinbettung, bei denen das letzte finite Verb fehlt, mitunter als grammatischer beurteilt werden, als die korrekten vollständigen Sätze (Gibson & Thomas, 1999). Dieses Ergebnis steht ebenfalls im Widerspruch zu den Vorhersagen, die man aus symbolischen Modellen ableiten würde (es sei denn, man nimmt zusätzliche Effekte aus Gedächtnisbeschränkungen an, vgl. Gibson & Thomas, 1999). Es deckt sich aber mit der Performanz der SRNs für das dritte Verb bei Zweifacheinbettung, auf dessen Position am stärksten die Unit aktiviert wird, die das Satzende anzeigt.

Die Ergebnisse zeigen eine hohe Übereinstimmung mit Befunden aus der psycholinguistischen Forschung. Bemerkenswert ist aber auch die Unerwartetheit sowohl der empirischen als auch der Simulationsergebnisse, wenn man sie aus der Perspektive symbolischer Modelle betrachtet, die eine unendliche grammatische Kompetenz annehmen. Christiansen und Chater (1999b) beziehen sich bei ihrem Versuch, die Ursachen und Mechanismen zu identifizieren, die den Verarbeitungsmerkmalen von SRNs zugrunde liegen, zunächst wiederum auf Gibson (1998). Im Sinne der Dependency Locality Theory (DLT) vermuten sie, dass die beschränkte repräsentationale Kapazität für den Leistungsabfall verantwortlich ist, der bei den Fällen auftritt, bei denen über längere Distanzen Dependenzinformationen aufrechterhalten werden müssen (etwa bei Zentraleinbettung gegenüber gekreuzter Abhängigkeit). Andererseits kritisieren sie, dass die Performanz sehr viele Merkmale aufweist, die nicht mit der DLT erklärt werden können – zum Beispiel die Tatsache, dass der Unterschied zwischen Zentraleinbettung und gekreuzter Abhängigkeit erst ab einer Einbettungstiefe von zwei Einbettungen auftaucht. Sie schließen daraus: „[...] the current results suggest that we should be wary of drawing strong conclusions for language processing behavior, in networks and perhaps also in people, from arguments concerning idealized infinite cases.“ (Christiansen & Chater, 1999b, S. 200). Diese Einschätzung setzt sich in der Annahme fort, dass Limitierungen der menschlichen Performanz (d.h. Verzögerungen, Fehlleistungen, Verstehensbeschränkungen, Ungenauigkeiten bei der Verarbeitung) nicht Resultate einer beschränkten Performanz über einer unbeschränkten Kompetenz sind, sondern aus intrinsischen Eigenschaften und Beschränkungen des Verarbeitungssystems resultieren.<sup>22</sup>

Klarzustellen ist hierbei, dass die Performanz von SRNs ab einer gewissen Einbettungstiefe gänzlich zusammenbricht – die Fähigkeit der Netzwerke zur Verarbeitung komplexer Strukturen bezieht sich ganz klar auf eine begrenzte Rekursivität. Da aber eben dieser Zusammenbruch der Verarbeitung auch für die menschliche Verarbeitung charakteristisch

---

<sup>22</sup>Damit kritisieren die Autoren auch symbolische Modelle menschlicher Sprachverarbeitung: Um Performanzeigenschaften zu generieren und zu erklären, müssen in diesen Modellen künstliche Beschränkungen angenommen und operationalisiert werden. Die hierfür eingeführten Beschränkungsfunktionen sind ein gewichtiger (und mitunter willkürlicher) freier Parameter, und stellen immer eine zusätzlich zu begründende Modellannahme dar. In SRNs hingegen treten die Beschränkungen automatisch als intrinsische Eigenschaft des Modells auf, die aus den architektonischen Beschränkungen und deren Interaktion mit den statistischen Eigenschaften des Inputs resultieren.

ist, ist dies durchaus ein erwünschter Effekt der Fehlerkorrespondenz zwischen Modell und empirischer Wirklichkeit. Aus dieser Perspektive wird die Kompetenz-Performanz-Dichotomie zur unrealistischen, idealisierten Modell-Annahme symbolischer Modelle (vgl. auch Seidenberg & MacDonald, 1999).

SRNs bieten also eine neue Perspektive zur Erklärung der Performanz für komplexe Satzstrukturen. Die genauen Vorhersagen hieraus sind allerdings mit Vorsicht zu genießen. Rohde (2002) formuliert eine allgemeine kritische Anmerkung zu Befunden aus SRNs und den daraus hervorgehenden empirischen Vorhersagen für Präferenzen in der Satzverarbeitung: Es muss in Frage gestellt werden, ob ein Modell des Sprachverstehens die gleichen Präferenzen zeigen würde wie ein Modell der Vorhersage syntaktischer Klassen. Ein Modell, das nur die Menge möglicher Folgewörter vorhersagen muss, hat den deutlichen Vorteil, dass es die dazu aktiv gehaltenen Informationen unmittelbar verwerfen kann, sobald diese für die Vorhersage irrelevant werden. Demgegenüber müsste ein Modell des Sprachverstehens alle relevanten semantischen Informationen mindestens bis zum Ende des Satzes aktiv halten, um eine kohärente und konsistente Interpretation des Satzes zu gewährleisten. Wird der Einfluss solcher Aspekte berücksichtigt, ändern sich die aus dem Modell abgeleiteten Präferenzen möglicherweise tiefgreifend. Aus dieser Perspektive sind SRNs sicherlich als eindimensionale Modelle zu verstehen, in denen man eine große Anzahl relevanter Informationen und Wechselwirkungen schlicht außer Acht lässt. Dieser Einschätzung stimmen auch Weckerly und Elman (1992) zu, wenn sie feststellen, dass lexikalische Antizipation nur einen Teil des Sprachverstehens ausmacht und eine schlechte Verstehensleistung schwerlich auf unpräzise Vorhersagen reduziert werden kann. Dennoch finden sich vermehrt empirische Hinweise darauf, dass die Antizipationsfähigkeit ein wichtiger Aspekt des Sprachverstehens ist und die Verstehensleistung mitbestimmt (vgl. u.a. Altmann, 1999; Altmann & Kamide, 1999; Konieczny, 2000; Konieczny & Döring, 2003; DeLong et al., 2005; Otten & van Berkum, 2008; T. F. Jaeger et al., submitted).

### 3.2.1.1. Exkurs: Performanz und globale Fehlermaße

Für die Analyse der Performanz von Netzwerken werden mathematische Distanz- bzw. Ähnlichkeitsmaße herangezogen. Dabei geht es in der Regel darum, den Output des Netzwerks mit externen Daten zu vergleichen, seien es die korrespondierenden Target-Muster, die dem Training zugrundegelegt wurden, oder andere Vergleichsdaten, wie zum Beispiel entsprechende Wahrscheinlichkeitsverteilungen aus Textkorpora oder absolute Übergangswahrscheinlichkeiten aus der probabilistischen kontextfreien Grammatik, aus der die Trainings- und Teststimuli erzeugt wurden. Die Nähe eines produzierten Output-Musters zu einem Target-Muster dient als Grundlage der Beurteilung der Performanz. Neben den Distanzmaßen kommen auch allgemeine Entropiemaße und elaboriertere Fehlermaße zum Einsatz. In der konnektionistischen Literatur im Bereich Sprachverarbeitung finden sich unter anderem folgende Maße: Euklidische Distanz, Cosinus, Wurzel der mittleren quadratischen Abweichung, Kullback-Leibler-Divergenz. Ein spezifischeres Maß des Fehlers, das als direktes Vergleichsmaß für empirische Effekte (Lesezeiten) herangezogen wird, ist der Grammatical Prediction Error (GPE, Christiansen & Chater, 1999b; MacDonald

& Christiansen, 2002).<sup>23</sup> Wir besprechen kurz die wichtigsten Maße und vergleichen sie anschließend.

Die euklidische Distanz bestimmt den geometrischen Abstand zweier Punkte im hochdimensionalen Zustandsraum. Über die Aktivierungswerte und Zielwerte der einzelnen Vektorstellen wird ihr Lokalisationspunkt in diesem theoretischen Raum erfasst. Der Abstand zwischen der Lokalisation der Vektoren  $V_{\text{output}}$  und  $V_{\text{target}}$  ist ein Maß für die Ähnlichkeit der Vektoren. Je kleiner die euklidische Distanz, desto ähnlicher sind die Vektoren. Die Distanz wird als Wurzel aus der Summe der quadrierten Differenzen der einzelnen korrespondierenden Vektorstellen berechnet (siehe Gleichung 3.1).

$$d(x, y) = \sqrt{\sum_{k=i}^n (x_i - y_i)^2} \quad (3.1)$$

Als weiteres Distanzmaß zweier Vektoren kann der Cosinus verwendet werden, der den Winkel zwischen den beiden Vektoren unabhängig von deren Länge wiedergibt. Je größer der Cosinus (dieser nimmt Werte zwischen 0 und 1 an), desto kleiner der Winkel und desto ähnlicher die Vektoren. Um daraus ein Fehlermaß zu erhalten, wird der Cosinus von einem Grundbetrag 1.0 abgezogen, so dass ein kleinerer Wert einer größeren Ähnlichkeit bzw. einem geringen Fehler entspricht. Im hochdimensionalen Vektorraum wird der Cosinus zweier Vektoren als Quotient des Skalarprodukts der Vektoren (Summe der Produkte der korrespondierenden Vektorstellen) durch das Produkt der Längen der beiden Vektoren (Produkt ihrer Beträge) bestimmt.<sup>24</sup> Das Fehlermaß  $1 - \text{COS}$  wird dementsprechend als Differenz dieses Quotienten von 1.0 berechnet (siehe Gleichung 3.2).

$$d(x, y) = 1 - \frac{\sum_{k=i}^n x_i \cdot y_i}{\sqrt{\sum_{k=i}^n x_i^2} \cdot \sqrt{\sum_{k=i}^n y_i^2}} \quad (3.2)$$

Die Wurzel der mittleren quadratischen Abweichung (Root Mean Square Error, RMSE) ist ein Maß zur Berechnung der Abweichung von einzelnen Datenpunkten von einem erwarteten Mittelwert für die Daten. Der RMSE kann als Schätzwert für die Abweichung zweier Vektoren zueinander verwendet werden. Er wird berechnet als Wurzel aus dem Quotienten der Summe der Residuen (Quadrate der Abweichung zwischen korrespondierenden Vektorstellen) und der Menge der Fälle (Vektorstellen). Dadurch werden die Differenzen der Werte der Vektoren als Distanz der quadratischen Mittelwerte in ein einzelnes Maß aggregiert (siehe Gleichung 3.3).

$$\text{RMSE}(x, y) = \sqrt{\frac{\sum_{i=1}^n (x_i - y_i)^2}{n}} \quad (3.3)$$

---

<sup>23</sup>Daneben existieren weitere komplexere Maße wie zum Beispiel die Simulated Reading Time (SRT, Rohde, 2002), die die Performanz verschiedener Verarbeitungskomponenten verrechnet (vgl. Kap. 4.2).

<sup>24</sup>Durch die Normierung über die Vektorlängen ist der Cosinus im Unterschied zur euklidischen Distanz unabhängig von der Länge der Vektoren, die Distanz wird alleine anhand des Winkels der Vektoren bestimmt.

Im Zusammenhang mit der Unterschiedlichkeit von Verteilungen tauchen auch die Begriffe relative Entropie (auch Kullback-Leibler-Entropie oder Kullback-Leibler-Divergenz) und Kreuzentropie auf. Entropie ist ein Begriff aus der Informationstheorie<sup>25</sup> und bezeichnet hier ein Maß für die Informationsdichte eines Zeichensystems. Im System wird der Informationsgehalt eines Zeichens über seine Auftretenswahrscheinlichkeit bestimmt – je geringer die Auftretenswahrscheinlichkeit eines Zeichens, desto größer der Informationsgehalt. Mit der Entropie eines Zeichens berechnet man den Erwartungswert seines Informationsgehalts innerhalb des Zeichensystems. Als Maß für die Unterschiedlichkeit zweier Wahrscheinlichkeitsverteilungen (bzw. als Qualitätsmaß zum Vergleich eines Modells einer Wahrscheinlichkeitsverteilung mit der beobachteten Wahrscheinlichkeitsverteilung) verwendet man sowohl die Relative Entropie als auch die Kreuzentropie. Die relative Entropie kommt auch in FFNs und SRNs zum Einsatz: Sie wird im Backpropagation-Algorithmus zur Bestimmung des Fehlers zwischen gewünschter und tatsächlicher Aktivierung verwendet. Sie kann aber auch dazu benutzt werden, um die Performanz eines Netzwerks zu bestimmen. Dabei gilt: Je geringer die Entropie, desto besser die Übereinstimmung zwischen Target-Vektor und Output-Vektor. Gleichung 3.4 gibt die Berechnung der relativen Entropie wieder. Die Kreuzentropie erzielt sehr ähnliche Ergebnisse wie die relative Entropie, ermöglicht unter bestimmten Voraussetzungen aber eine Vereinfachung der Berechnung. Sie wird berechnet als Summe aus Entropie und Relativer Entropie (vgl. Gleichung 3.5).

$$d(x|y) = \sum_i x_i \cdot \log_2 \frac{x_i}{y_i} \quad (3.4)$$

$$d(x|y) = \sum_i x_i \cdot \log_2 \left( \frac{x_i}{y_i} \right) + (1 - x_i) \cdot \log_2 \left( \frac{1 - x_i}{1 - y_i} \right) \quad (3.5)$$

Ein sehr viel spezifischeres Maß für die Untersuchung syntaktischer Aspekte ist der Grammatical Prediction Error (GPE, nach Christiansen & Chater, 1999b; MacDonald & Christiansen, 2002).<sup>26</sup> Dieser vergleicht den Unterschied zwischen der Output-Aktivierung und den absoluten Wortübergangswahrscheinlichkeiten und berücksichtigt beim Vergleich der Vektoren neben der Übereinstimmung von Aktivierung auch die negative fehlende Übereinstimmung (d.h. die Aktivierung von Units, die nicht aktiviert sein sollten). Er dient damit als spezielles Maß, mit dem die Grammatikalität des Outputs erfasst werden kann: Dabei wird der Output detailliert mit der Verteilung der Wortklassenwahrscheinlichkeit verglichen, die für das entsprechende Folgeelement einer Sequenz grammatikalisch erwartet wird. Der GPE wird als Differenz zwischen dem Grundbetrag 1.0 und dem Quotienten der *hits* und der Summe aus *hits*, *false alarms* und *misses* berechnet (siehe Gleichung 3.6). Die *hits* entsprechen dabei der Summe der im Vergleich zum Wahrscheinlichkeitsvektor korrekt aktivierten Units, *false alarms* sind die Summe der falsch aktivierten Units (Units, die nicht aktiviert sein sollten), *misses* sind die Summe der fehlenden Aktivierung von Units (Units, die stärker aktiviert sein sollten). *False alarms* und *misses* entsprechen also zusammen der Summe der Beträge, die sich aus den Differenzen zwischen den Vektorstellen ergeben. Die

<sup>25</sup>Er wird in Analogie zur Verwendung des Begriffs in der Physik (speziell der Thermodynamik) verwendet, und geht im informationstheoretischen Zusammenhang auf C.E. Shannon zurück.

<sup>26</sup>Eine ähnliche Berechnung sieht der GPE bei Van der Velde et al. (2004) vor.

Gesamtaktivierung der Units, die im Divisor den relevanten Anteil für die Bemessung der Aktivierungspräzision stellt, ergibt sich allerdings bereits aus der Summe von *hits* und *false alarms* – aus welchem Grund die *misses* in den Divisor aufgenommen wurden ist hingegen nicht nachvollziehbar. Der GPE ist als verbessertes Vergleichsmaß für Performanzdaten aus empirischen Studien konzipiert. Unter anderem wird für ihn beansprucht, eine gute Annäherung an Lesezeiten und Grammatikalitätsratings zu sein.

$$\text{GPE} = 1 - \frac{\text{hits}}{\text{hits} + \text{false alarms} + \text{misses}} \quad (3.6)$$

**Problematisierung globaler Fehlermaße** Die hier vorgestellten Maße haben verschiedene Vor- und Nachteile bei der Verwendung als Qualitätsmaß der Performanz. Unter anderem zeigen sie verschiedene Grade an Sensitivität für Merkmale der Differenzen zwischen den Vektoren. So misst der Cosinus nur den Winkel zwischen den Vektoren, nicht aber die Differenz der Strecken der Vektoren. Durch die Verwendung des Produkts der Vektorstellen werden beim Cosinus nur die Aktivierungen berücksichtigt, die im Target-Vektor erwartet werden. Falsch aktivierte Units gehen (da mit 0 multipliziert) nicht in die Berechnung mit ein. Diese Eigenschaft teilt der Cosinus mit den Entropiemaßen, die den Logarithmus des Quotienten der Vektorstelle aus der Aktivierung und dem Target verwenden, und folglich ebenfalls nur die Vektorstellen mit Target-Aktivierung berücksichtigen und falsch aktivierte Units nicht bestrafen.<sup>27</sup> In dieser Hinsicht scheinen Maße, die Beträge der Vektorstellen-Differenzen verwenden, besser geeignet. Das Problem der stärkeren Gewichtung von Ausreißern beim Root Mean Square Error durch die Quadrierung der Differenzen der Vektorstellen ist hingegen weniger problematisch: In der Regel werden die Aktivierungen und Targetvektoren normalisiert, und die Quadrierung führt zu einer höheren Sensitivität der Methode für abweichende Aktivierungen. Die Normalisierung gleicht auch eine Problematik der euklidischen Distanz aus, die sehr stark durch unterschiedliche Skalen der Variablen (hier: Vektorstellen) beeinflusst werden kann, d.h. bei nicht normalisierten Vektorstellen können einzelne Werte die gemessene Distanz stark beeinflussen. Allerdings berücksichtigen Distanzmaße wie die die Euklidische Distanz und in dieser Form alle Maße, die auf Differenzen basieren nur die Vektorstellen, die eine Differenz aufweisen. Abweichende Werte werden also nicht (bzw. nur indirekt) in Relation gesetzt zur Menge der übereinstimmenden Vektorstellen. Dieses Vorgehen finden wir ausschließlich beim GPE.

Ein weiteres Problem der vorgestellten globalen Maße – bezogen auf die Beurteilung der kombinatorischen Systematizität der Generalisierung in solchen sequenzverarbeitenden Netzwerken – wird von S. L. Frank (2006b) thematisiert: Den Fehlermaßen fehlt eine Baseline, ab der für die Performanz behauptet werden kann, besser zu sein als in einem nicht-generalisierenden (d.h. memorierenden) Netzwerk. Für seine Studie entwirft er deshalb ein Performanz-Maß mit Baseline (Gleichung 3.7). In diesem wird zunächst der Anteil grammatischer Aktivierungen ( $a(G)$ ) berechnet: Werden ausschließlich Units

---

<sup>27</sup>In diesem Zusammenhang wird Entropie auch als Grundlage für die Fehlerbestimmung für die Gewichts-anpassung in SRNs verwendet, da falsche Aktivierungen nicht berücksichtigt werden und nur positive Übereinstimmungen zu einer Gewichts-anpassung führen. In diesem Zusammenhang spricht man von Verstärkungslernen.

aktiviert, die Wörter einer korrekten grammatischen Fortsetzung repräsentieren, ist  $a(G) = 1$  und die Performanz erhält den Wert +1. Werden ausschließlich Units ungrammatischer Fortsetzungen aktiviert bzw. alle grammatischen Fortsetzungen nicht aktiviert, ist  $a(G) = 0$  und die Performanz erhält den Wert -1. Den Baseline-Wert von 0 erhält die Performanz über den Vergleich mit der Aktivierung von Units nach Bigramm-Wahrscheinlichkeit.<sup>28</sup> Diese Wahrscheinlichkeit  $p(G)$  wird aus der Verteilung in den Trainingssätzen abgeleitet, und unterscheidet sich von der Wahrscheinlichkeit des Folgeworts, wenn die Wörter der vorausgegangenen Sequenz berücksichtigt werden. Wenn nur die Leistung eines solchen nicht-generalisierenden Netzwerks erreicht und folglich die Aktivierung  $a(G) = p(G)$  erzeugt wird, erhält die Performanz den Wert 0.<sup>29</sup> Aus diesen Festlegungen errechnet sich die Performanz in Gleichung 3.7.

$$\text{Performanz} = \begin{cases} \frac{a(G) - p(G)}{p(G)}, & \text{für } a(G) \leq p(G) \\ \frac{a(G) - p(G)}{1 - p(G)}, & \text{für } a(G) > p(G) \end{cases} \quad (3.7)$$

Abgesehen von allgemeinen Schwächen verschiedener Maße sollte ein allgemeines Problem globaler Fehlermaße (einschließlich des GPE) berücksichtigt werden: Das Problem, dass es eben globale Fehlermaße sind. In der Praxis der Analyse konnektionistischer Netzwerke hat sich wiederholt gezeigt, dass die ausschließliche Berücksichtigung des globalen Fehlers Fehlinterpretationen begünstigt. Diese beruhen darauf, dass der Fehler allein noch keine Aussagen über seine Ursachen zulässt. In einer Modellierungsstudie überprüfen Konieczny, Ruh und Müller (2009) die Simulationsergebnisse von MacDonald und Christiansen (2002) zu Verarbeitungsunterschieden für verschiedene Relativsatzarten und Unterschiede des Trainingsumfangs bzw. der Erfahrungheit (vgl. folgende Unterkapitel). Wie sich hier zeigte, sind die Unterschiede im globalen Fehlermaß nicht auf graduelle Unterschiede der Aktivierung korrekter Units zurückzuführen, sondern entstehen durch die Aktivierung völlig falscher Units. Diese Fehlaktivierung beruht auf einer unausgereiften Kategorisierung der Wortklassen. Sie weist auf ein unvollständiges Training der Netzwerke hin und ist also besser als Artefakt des Simulationsverfahrens zu verstehen. Um der tatsächlichen Performanz und der Arbeitsweise der Netzwerke (einschließlich ihrer Beschränkungen) auf den Grund zu gehen, erweist es sich als unabdingbar, zur Analyse die Aktivierungsmuster für Detailuntersuchungen heranzuziehen, um die tatsächliche Quelle von Fehlern identifizieren und dadurch spezifischere Aussagen über die Performanz treffen zu können (vgl. auch Konieczny & Müller, 2010; Weldle, 2005; Konieczny & Ruh, 2003).

Abschließend wollen wir noch ein Verfahren erwähnen, das die Performanz anhand der Analyse interner Repräsentationen bemisst: Tabor, Juliano und Tanenhaus (1997) stellen mit dem Visitation Set Gravitation (VSG) Modell eine Methode zur Ableitung

<sup>28</sup>Dies entspricht der Wortfolgewardrscheinlichkeit, wenn ausschließlich das aktuelle Wort berücksichtigt wird, und die Constraints der Wörter aus der vorausgegangenen Sequenz unberücksichtigt bleiben. Diese Performanz würden perfekt trainierte FFNs für die Wortvorhersage-Aufgabe erzeugen.

<sup>29</sup>Das bedeutet, dass alle Satzpositionen, auf denen  $p(G) = 1$  ist, nicht mit diesem Score erfasst werden können, denn hier könnte selbst bei perfekter Generalisierung nur ein Performanz-Score von 0 erzielt werden. Dies ist in der Studie von S. L. Frank (2006b) zum Beispiel nach einem EOS-Marker (end-of-sentence) der Fall.

von Performanzdaten aus den internen Zuständen des Netzwerks vor, die zum Vergleich mit empirischen Daten herangezogen werden können (vgl. auch Tabor & Tanenhaus, 1999). Grundlage ist ein einfaches SRN mit Wortvorhersage-Aufgabe, dessen Hidden-Layer-Aktivierungen während der Verarbeitung aufgezeichnet und im hochdimensionalen Vektorraum verortet werden. Diese Verortungen entsprechen fixierten Massepunkten, die eine Gravitationskraft ausüben. Cluster solcher Massepunkte entsprechen Attraktoren. Die Ableitung von Lesezeiten erfolgt anschließend im Gravitationsmodul, einem dynamischen Clusterungs-Mechanismus: Um die Lesezeit eines Wortes in Abhängigkeit seines syntaktischen Kontexts abzuleiten, wird die Verortung seiner Hidden-Layer-Aktivierung als neuer Massepunkt behandelt, der sich frei zwischen den fixierten Massepunkten bewegen kann. Nun wird die Zeitspanne bemessen, die der frei bewegliche Massepunkt benötigt, um in einem der Attraktoren einen stabilen Zustand zu finden. Die benötigte Zeitspanne stellt eine Schätzung der erwarteten Lesezeit für das Wort im entsprechenden Satz dar. Die Attraktoren im hochdimensionalen Raum können als erfahrungsbasierte Prototypen von Wortklassen aufgefasst werden. Damit implementiert das Modell einen gebrauchsbasierten Ansatz der Sprachverarbeitung (Tabor & Tanenhaus, 1999, S. 494):

Thus, the gravitational mechanism implements the idea that, in responding to a new instance of a word-in-context, the processor analogizes that word-in-context to its previous experiences and gravitates to a cluster corresponding to the most-similar previous experience.

### 3.2.2. Frequenz, Regularität, Erfahrung

In vielen Studien wird auf die zentrale Bedeutung der Verteilungsmerkmale des Inputs und die Sensitivität konnektionistischer Modelle für die Input-Strukturen hingewiesen. Dieser Bezug rückt konnektionistische Modelle in die Nähe probabilistischer Sprachverarbeitungstheorien. Dabei spielen auch Lernmechanismen und Repräsentationsstrukturen der Modelle eine wichtige Rolle, über die der Einfluss der Input-Eigenschaften und der Erfahrung mit diesen erst zum Tragen kommen. Zur Erklärung von Verhaltensmustern wird die Interaktion von Frequenz und Regularität als zentrales Prinzip herangezogen<sup>30</sup>, das die internen Repräsentationen beeinflusst und damit auf das Verarbeitungsverhalten wirkt. Dabei haben Erfahrungsunterschiede (hier: Dauer des Umgangs mit dem Input bzw. Anzahl der Epochen im Training) einen deutlichen Einfluss auf die Ausprägung der Interaktion und ihre Auswirkungen.

**Ressourcenbeschränkung vs. Erfahrung** Bezogen auf die Verarbeitung verschiedener syntaktischer Strukturen und auf Verarbeitungseigenschaften von SRNs als Modell der menschlichen Sprachfähigkeit diskutieren MacDonald und Christiansen (2002) die Frage, durch welche Aspekte unseres kognitiven Systems unsere Verarbeitungskapazitäten bestimmt werden und wodurch in diesem System Komplexität, Verarbeitungspräferenzen und interindividuelle Verarbeitungsunterschiede bedingt sind. Im Zentrum steht eine konnektionistische Konzeption der Sprachverarbeitung, in der die Verarbeitungskapazität aus

---

<sup>30</sup>Hier ist anzumerken, dass die Interaktion von Frequenz und Regularität auch schon in früheren konnektionistischen Ansätzen zur Sprachverarbeitung als elementarer Aspekt zur Erklärung der Effekte herangezogen wurde (vgl. z.B. Seidenberg & McClelland, 1989; Hare & Elman, 1995).

der Netzwerk-Architektur und der Erfahrung mit den zu verarbeitenden Strukturen entsteht. Die Kapazität als kognitive Ressource ist in dieser Konzeption nicht mehr nur als Primitiv zu erfassen, das beliebig und unabhängig vom sprachverarbeitenden System variieren kann. Daraus wird abgeleitet, dass interindividuelle Unterschiede beim Satzverstehen nicht aus Variationen einer separaten Arbeitsgedächtniskapazität stammen – vielmehr entwickeln sich diese Unterschiede in der Interaktion biologischer Faktoren mit Spracherfahrung.

MacDonald und Christiansen (2002) stellen mit ihrem Ansatz einen konnektionistischen Gegenentwurf zu prominenten Ansätzen der Sprachverarbeitung vor, in denen Performanzaspekte unabhängig vom sprachverarbeitenden System und dessen Repräsentationen vorwiegend durch Kapazitätsbeschränkungen des Arbeitsgedächtnisses bestimmt werden. Speziell beziehen sie sich dabei auf den Ansatz von Just und Carpenter (1992) und den kritischen Gegenentwurf dazu von Waters und Caplan (1996). Während diesen Ansätzen jeweils unterschiedliche Annahmen über die Natur des Arbeitsgedächtnisses zugrunde liegen<sup>31</sup>, teilen sie die Annahme der Trennung von sprachlichem Wissen (und dessen Repräsentationen) und dem Arbeitsgedächtnis. Beide vermuten den Ursprung interindividueller Unterschiede in Kapazitätsunterschieden sowie in der Interaktion der getrennten Komponenten Wissen und Verarbeitung. MacDonald und Christiansen (2002) kritisieren diese Trennung von Wissen und Verarbeitung bzw. Verarbeitungskapazität.<sup>32</sup> Sie stellen in diesem Zusammenhang auch in Frage, was die Tests zur Ermittlung interindividueller Unterschiede der Kapazität des Arbeitsgedächtnisses (v.a. der Reading-Span-Aufgabe, Daneman & Carpenter, 1980)<sup>33</sup>, die den Evaluationen empirischer Ergebnisse von Just und Carpenter (1992) sowie Waters und Caplan (1996) in verschiedener Form zugrunde liegen, eigentlich aussagen.

<sup>31</sup>Ein zentraler Unterschied zwischen den Ansätzen beruht in den Annahmen zum Arbeitsgedächtnis, dessen Arbeitsweise und dessen Interaktion mit den sprachbezogenen Verarbeitungskomponenten. Die Debatte zwischen den beiden Ansätzen kulminiert in ihrer unterschiedlichen Beurteilung dessen, wodurch interindividuelle Unterschiede in der Sprachverarbeitung bedingt sind. Just und Carpenter (1992) nehmen in ihrer *Capacity Theory of Comprehension* an, dass das Arbeitsgedächtnis eine einzige Ressource ist, aus der Aktivierung für alle kognitiven Prozesse geschöpft wird. Verarbeitungsunterschiede entstehen aus der unterschiedlichen Spanne des Arbeitsgedächtnisses: Je größer die Spanne ist, desto mehr Informationen können gleichzeitig zur korrekten Verarbeitung herangezogen werden, und desto mehr Interpretationsalternativen können gleichzeitig aufrechterhalten und verfolgt werden. Sehr unterschiedliche Effekte in verschiedenen sprachlichen Verarbeitungsbereichen können also auf die gleiche Ursache zurückgeführt werden. Waters und Caplan (1996) hingegen postulieren die Unterteilung des Arbeitsgedächtnisses in mindestens zwei spezialisierte Submodule: Ein Modul ist für die Verarbeitung unbewusst ablaufender automatisierter Prozesse zuständig, die zum Beispiel bei der Zuweisung syntaktischer Struktur ablaufen. Ein davon zu unterscheidendes zweites Modul steuert die bewussten Prozesse, die auf kontrollierte Weise erfolgen. Für unterschiedliche Effekte in verschiedenen sprachlichen Verarbeitungsbereichen muss deshalb berücksichtigt werden, dass sie auf einem Zugriff auf unterschiedliche Ressourcen beruhen. Interindividuelle Unterschiede beruhen in diesem Ansatz also auf Unterschieden verschiedener sprachspezifischer Module und verschiedener Ressourcen (spezialisierte Arbeitsgedächtnis-Komponenten).

<sup>32</sup>Mit ihrer Kritik an der Trennung von sprachlichem Wissen und sprachunabhängigen Performanzaspekten eines gekapselten Arbeitsgedächtnisses wird auch indirekt wieder die Kompetenz-Performanz-Dichotomie infrage gestellt (vgl. Christiansen & Chater, 1999b).

<sup>33</sup>Die Reading-Span-Aufgabe von Daneman und Carpenter (1980) bemisst die Kapazität des sprachlichen Arbeitsgedächtnisses anhand einer kombinierten Aufgabe, in der eine steigende Anzahl von Sätzen gelesen und das letzte Wort des Satzes für eine spätere Abfrage gemerkt werden muss. Anhand der Leistung in diesem Test können Versuchsteilnehmer (VTn) in verschiedene Leistungsgruppen – sogenannte low-, mid- und high-span readers – eingeteilt werden.

**Effekte von Regularität und Erfahrung** Werfen wir einen Blick auf die Simulationsstudie von MacDonald und Christiansen (2002). Zur Untersuchung greifen sie einen der Gegenstände von King und Just (1991) auf: den Komplexitätsunterschied zwischen Objekt-Relativsätzen (Bsp. 3c) und Subjekt-Relativsätzen (3b) und die Verarbeitungsunterschiede für diese Satzstrukturen in Abhängigkeit interindividueller Unterschiede. Für diese Satztypen wurden in einer Vielzahl von Studien Komplexitätsunterschiede festgestellt, die eine höhere Komplexität für Objekt-Relativsätze ausweisen (für eine Übersicht vgl. Gennari & MacDonald, 2008). King und Just (1991) untersuchen in ihrer Studie anhand von Lesezeitmessungen den Einfluss unterschiedlicher Arbeitsgedächtniskapazitäten für Sprachverarbeitungsprozesse. Sie stellen dabei eine Interaktion zwischen Kapazität und Satztyp fest: Kapazitätsunterschiede spielen nur bei der Performanz für Objekt-Relativsätze eine Rolle, nicht aber für Subjekt-Relativsätze. Dieser Unterschied wurde als Vorteil einer höheren Arbeitsgedächtnis-Kapazität für komplexere Satzstrukturen interpretiert, die für die einfacheren Subjekt-Relativsätze noch nicht ausgeschöpft werden muss (vgl. auch Just & Carpenter, 1992).

In ihren Simulationen mit SRNs vergleichen MacDonald und Christiansen (2002) die Performanz der Netzwerke für die verschiedenen Satztypen anhand des globalen Fehlers GPE auf den kritischen Satzpositionen. Zugleich untersuchen sie auch Performanzunterschiede in Abhängigkeit von der Anzahl trainierter Epochen. Aus den Simulationen ergeben sich zwei zentrale Befunde: (a) Tatsächlich finden sich Verarbeitungsunterschiede zwischen Relativsätzen mit Objekt- und Subjektivextraktion, die auf eine höhere Komplexität von Objekt-Relativsätzen hinweisen. (b) Ebenfalls findet sich eine Interaktion zwischen den Relativsatztypen und der Trainingsdauer, nach der sich nur die Performanz für Objekt-Relativsätze mit fortgesetztem Training verbessert, während sie für SRNs bereits von Beginn an vergleichsweise gut ist. In ihrer Interpretation setzen die Autoren Differenzen der Trainingsdauer der SRNs mit den unterschiedlichen Kapazitätsgruppen aus der empirischen Studie gleich. Die beiden Ergebnisse führen zu einem Verhältnis der Performanz in den unterschiedlichen Bedingungen, das eine gewisse Ähnlichkeit zu den korrespondierenden Verhältnissen der empirischen Studie von King und Just (1991) aufweist.<sup>34</sup>

MacDonald und Christiansen (2002) erklären die Komplexitätsunterschiede zwischen Objekt- und Subjekt-Relativsätzen als Effekt der unterschiedlichen Regularität der beiden Satzstrukturen. Im Englischen herrscht in einfachen Hauptsätzen die Satzstruktur  $NP_{\text{Subjekt}}-V_{\text{finit}}-NP_{\text{Objekt}}$  vor. Subjekt-Relativsätze spiegeln diese Abfolge wider und weisen damit eine höhere Regelmäßigkeit auf als Objekt-Relativsätze, die die seltene unregelmäßige Satzstruktur  $NP_{\text{Objekt}}-NP_{\text{Subjekt}}-V_{\text{finit}}$  besitzen. Dieser Unterschied schlägt sich direkt auf die Wortvorhersage-Aufgabe nieder, die die SRNs erfüllen müssen: Das strukturelle Wissen für einfache Hauptsätze kann direkt auf die Verarbeitung von Subjekt-Relativsätzen übertragen werden, innerhalb denen die korrekte Vorhersage für Folgewortklassen dadurch vergleichsweise einfach ausfällt. Demgegenüber muss die unregelmäßige Struktur von Objekt-Relativsätzen erst (als grammatikalisch erkannt und) erworben werden, bevor innerhalb dieses syntaktischen Kontexts korrekte Wortfolge-Vorhersagen getroffen werden können.

---

<sup>34</sup>Für eine kritische Analyse der Ergebnisse vgl. Konieczny, Ruh und Müller (2009); Konieczny und Müller (2010), siehe auch die Debatte im folgenden Kapitel 3.2.3.

**Interaktion von Frequenz und Regularität** Die Erklärung interindividueller Unterschiede erfolgt in Erweiterung dieser Argumentation anhand der Interaktion von Frequenz und Regularität: Da in Subjekt-Relativsätzen die regelmäßige Wortreihenfolge einfacher Hauptsätze aufgenommen wird, können sie durch den Transfer von einfacher auf die komplexe Struktur bereits früh bzw. mit wenig Erfahrung erfolgreich verarbeitet werden. Deshalb ändert sich die Performanz für diese Satzstruktur durch wachsende Erfahrung kaum. Aufgrund der Seltenheit von Relativsätzen im Vergleich zu anderen Satzstrukturen muss die unregelmäßige Satzstruktur von Objekt-Relativsätzen hingegen sehr häufig gesehen oder gehört worden sein, damit sie korrekt verarbeitet werden kann. Deshalb zeigt sich für diese Satzstruktur in der Simulation über mehrere Epochen bzw. mit wachsender Erfahrung ein deutlicher Effekt für die Qualität der Performanz. Als Gegenkonzept zu ressourcenbasierten Sprachverarbeitungstheorien, die interindividuelle Unterschiede auf Kapazitätsunterschiede zurückführen, werden aus dieser konnektionistischen Perspektive also Erfahrungsunterschiede als Erklärungsansatz für die Unterschiede herangezogen (MacDonald & Christiansen, 2002, S. 44):

Whereas comprehenders with limited experience can rapidly compute simple constraints to resolve ambiguities, only more experienced comprehenders have encountered complex combinatorial constraints often enough to compute them efficiently.

Mit dieser Reinterpretation des Ursprungs interindividueller Unterschiede stellen sie auch in Frage, was mit Kapazitätstests wie der Reading Span-Aufgabe überhaupt gemessen wird. Sie nehmen an, dass diese nur allgemeine sprachliche Erfahrungsunterschiede widerspiegeln. Darüber hinaus bestreiten sie aber nicht, dass Erfahrungsunterschiede nicht alle Unterschiede erklären können und es auch biologische Faktoren geben muss, die die Interaktion zwischen kognitivem System und sprachlicher Umwelt prägen.<sup>35</sup>

Die Ergebnisse von (MacDonald & Christiansen, 2002) wurden auch zu Multiple-Constraint-Ansätzen<sup>36</sup> und zu Ansätzen des statistischen Sprachlernens bzw. der erfahrungsbasierten Sprachverarbeitung in Beziehung gesetzt. Konkret auf die Gegenstände der Simulationsstudie bezogen untersuchen Wells et al. (2009) den Einfluss der Erfahrung im Sinne statistischen Lernens auf die Performanz bei der Verarbeitung von Relativsätzen. Die Versuchsteilnehmer (VTn) wurden hierzu in Gruppen eingeteilt, die mit komplexem Lesematerial arbeiten mussten, das entweder Relativsätze oder Komplementsätze enthielt. Nach längerer Lernzeit wurden die Gruppen auf Unterschiede bei der Verarbeitung von Relativsätzen getestet. Der Erfahrungsunterschied für diese Zielstruktur schlug sich korrespondierend zu den Simulationsergebnissen in der Performanz der VTn nieder: Zum einen

<sup>35</sup>MacDonald und Christiansen (2002) identifizieren verschiedene Eigenschaften des Systems, für die sie eine bedeutsame Rolle für die Interaktion mit Erfahrungseffekten annehmen, wobei Unterschiede der Präzision der phonetischen Repräsentation eine zentrale Rolle spielen.

<sup>36</sup>Im Rahmen theoretischer Ansätze unter diesem Label können SRNs als Implementierung bzw. als Werkzeug zur Erforschung von Constraint-Satisfaction-Prozessen in der menschlichen Sprachverarbeitung und zur Konkretisierung von Annahmen innerhalb dieses theoretischen Rahmens verstanden werden. Kern der Annahme von Multiple-Constraint-Ansätzen sind hoch interaktive Verarbeitungsprozesse, die zur probabilistischen Auflösung ambiger Strukturen alle zur Verfügung stehenden kontextuellen Informationen verschiedener Quellen parallel nutzen, um das Feld alternativer Interpretationen zu bewerten oder zu reduzieren. Ein Schwerpunkt liegt dabei auf der Untersuchung, wie Menschen probabilistische Informationen aus ihrer Umwelt und deren Interaktionen erwerben, repräsentieren und benutzen (vgl. z.B. MacDonald & Seidenberg, 2006).

fürte die Erfahrung zu geringeren Latenzen in den Lesezeiten auf den kritischen Positionen. Zum anderen war der Vorteil asymmetrisch für Subjekt- und Objekt-Relativsätze insofern er sich wesentlich stärker auf zweite auswirkte. Ähnliche Beziehungen zwischen sprachlicher Erfahrung und der Performanz bei der Verarbeitung von Objekt-Relativsätzen wurden zum Beispiel von Gennari und MacDonald (2008) untersucht und in Beziehung zu Verteilungsstrukturen gesetzt, die aus Präferenzen der Sprachproduktion entstammen – im Sinne einer komplexen Interaktion zwischen probabilistischen Eigenschaften der sprachlichen Struktur mit Produktions- und Verarbeitungs-Präferenzen (vgl. Gennari & MacDonald, 2009). Speziell zur Frage der interindividuellen Unterschiede liefern Misyak, Christiansen und Tomblin (2008) weitere Befunde aus Experimenten in den Paradigmen Artificial Grammar Learning und Serial Reaction Time. Sie untersuchen Eigenschaften von VTn beim statistischen Lernen von Fernabhängigkeiten in Kunstsprachen, bei denen die Menge des intervenierenden Materials variiert wurde. Die Befunde wurden mit der Performanz bei der Verarbeitung komplexer syntaktischer Strukturen sowie mit Befunden aus einer SRN-Modellierungsstudie verglichen. Misyak et al. (2008) schließen aus ihren Ergebnissen, dass Eigenschaften des statistischen Lernens eine zentrale Rolle für Komplexitätsunterschiede und interindividuelle Unterschiede bei der Satzverarbeitungsleistung spielen, und weitergehend, dass allgemeine Fähigkeiten bei statistischen Lernaufgaben ein besserer Prediktor für interindividuelle Unterschiede sind als Tests zur Spanne des verbalen Arbeitsgedächtnisses.

Um die genannten Befunde nochmals zusammenzufassen: Zentraler Faktor, der die Performanz von SRNs bestimmt und der in der Folge als Erklärung für menschliche Verarbeitungseigenschaften herangezogen wird, ist die Interaktion von Frequenz und Regularität sprachlicher Stimuli. Abschließend soll noch erwähnt werden, dass die daraus abgeleitete unterschiedliche Lernbarkeit sprachlicher Strukturen auch als konnektionistischer Erklärungsansatz für Effekte des Sprachwandels herangezogen wurde. Auf dieser Basis erklären Hare und Elman (1995) in einer Modellierung des morphologischen Sprachwandels mit einfachen FFNs (vgl. Kap. 2.2.1.1) die sprachgeschichtliche Entwicklung der englischen Flektionsklassen. Sie erweitern damit das Argument von Rumelhart und McClelland (1986a), dass Erwerb und Verarbeitung verschiedener Flektionsklassen innerhalb eines Systems erfolgen, und nicht wie in klassisch-symbolischen Ansätzen angenommen über getrennte Module für regelmäßige und irreguläre Flektionsformen (vgl. auch McClelland & Patterson, 2002; Pinker & Ullman, 2002). In ihrer Erweiterung fordern sie, dass anhand des angenommenen Systems auch Effekte des Sprachwandels berücksichtigt werden können müssen. Zentrales Argument von Hare und Elman (1995) ist die unterschiedliche Lernbarkeit (hinsichtlich Lerngeschwindigkeit und Fehlerrate) verschiedener Types und Tokens sowie Tendenzen zur Regularisierung. Die Lernbarkeit über mehrere seriell geschaltete Generationen von FFNs wird durch die Frequenz und die morphologische Regularität der Types und Tokens bestimmt. Ihre Interaktion beeinflusst maßgeblich die Übergänge von Items zwischen verschiedenen Klassen und die Herausbildung neuer übergeordneter Klassen. Ähnliche Lernbarkeitsargumente wurden in weiteren konnektionistischen Studien zur Typologie für die Erklärung sprachlicher Universalien und zur Erklärung von Veränderungsmustern beim Sprachwandel herangezogen (vgl. Christiansen & Chater, 2008; Bybee & McClelland, 2005; Weldle, 2005; Christiansen et al., 2002; van Everbroeck, 2003; S. Levy et al., 2000; Christiansen & Devlin, 1997).

### 3.2.3. Sensitivität für den lokalen Kontext

Konieczny, Ruh und Müller (2009) äußern eine detaillierte Kritik an der methodischen Vorgehensweise und der Interpretation der Daten in den Simulationen von MacDonald und Christiansen (2002). In ihren Replikationen stellen sie fest, dass der zentrale Befund zum Ursprung syntaktischer Komplexität auf Artefakten beruhen (vgl. auch Konieczny & Müller, 2010; Konieczny & Ruh, 2003). MacDonald und Christiansen (2002) nehmen an, dass die höhere Komplexität von Objekt-Relativsätzen auf der seltenen unregelmäßigen Wortfolgestruktur beruht. Konieczny und Kollegen stellen dieser Interpretation der Simulationsresultate die Annahme entgegen, dass der höhere Anteil an Fehlaktivierungen für Objekt-Relativsätze auf den unausgereiften internen Repräsentationen der SRNs nach zu kurzem Training beruht, aufgrund derer falsche Wortvorhersagen produziert werden. Dabei kritisieren die Autoren aber nicht prinzipiell den Erklärungsansatz mit SRNs oder die Aufhebung der Trennung zwischen Arbeitsgedächtnis und sprachlichem Wissen. Vielmehr geht es um eine tiefergehende und besser begründete Erklärung der Verarbeitungsvorgänge in SRNs und um eine besser reflektierte Adaption dieser Vorgänge zur Erklärung empirischer Phänomene und den kognitiven Mechanismen, die diesen zugrunde liegen. Um die folgende Diskussion nachvollziehbar zu gestalten, stellen wir den Gegenstand in Beispiel 4 mit markierten kritischen Positionen dar.

- (4) a. ORS: The reporter that<sub>Relativpronomen</sub> the senator attacked<sub>Relativsatz-Verb</sub> admitted<sub>Matrix-Verb</sub> the error eos
- b. SRS: The reporter that<sub>Relativpronomen</sub> attacked<sub>Relativsatz-Verb</sub> the senator admitted<sub>Matrix-Verb</sub> the error eos

**Detailanalyse des globalen Fehlers** Die Kritik und Neuinterpretation der Simulationen von MacDonald und Christiansen (2002) basiert auf der Problematisierung der dortigen Interpretation des Netzwerkverhaltens auf Grundlage des globalen Fehlers (GPE): Dessen Ausprägungen lassen keinen genauen Rückschluss auf die Ursachen des Fehlers zu. Die im Folgenden erläuterten tatsächlichen Fehlerursachen besitzen eine wesentlich höhere Aussagekraft und ermöglichen einen direkten Zugang zu den Verarbeitungseigenschaften von SRNs. In ihrer kritischen Replizierung verfolgen Konieczny, Ruh und Müller (2009) die Frage „What’s in an Error?“ (Titel der Studie): Was passiert eigentlich genau auf den fehlerbelasteten Stellen? Welche Komponenten des globalen Fehlers GPE sind für den hohen Fehler auf den kritischen Positionen verantwortlich – d.h. entstammt der Fehler eher der Aktivierung falscher Units oder der zu geringen Aktivierung korrekter Units? Und vor allem: Welche Wortklassen werden tatsächlich im Detail vorhergesagt? Um diese Fragen zu beantworten, untersuchen die Autoren zunächst die positionsabhängigen Output-Aktivierungen für die kritischen Positionen (vergleichbar mit der Vorgehensweise von Elman, 1991). Von besonderem Interesse waren dabei die Positionen, auf denen der größte Anteil der erfahrungsbasierten Unterschiede zwischen Subjekt-Relativsätzen (SRS) und Objekt-Relativsätzen (ORS) zu Tage trat (vgl. Bsp. 4b und 4a), und auf denen in der Folge die Interpretation als Interaktion aus Frequenz und Regularität begründet wurde. Zusätzlich wurden mit dem Verfahren der Multidimensionalen Skallierung (vgl. Kap. 3.1.3.3) Analysen der internen Repräsentationen zu verschiedenen Trainingsepochen

durchgeführt, um weitere Anhaltspunkte zur Klärung der Performanz-Merkmale zu finden. Ihre Analysen ergaben folgende Ergebnisse:

- Position des finiten Relativsatz-Verbs in ORS (in Bsp. 4a: *attacked*). Hier entstammt der Fehler zwei Quellen: Zum einen aus der starken Aktivierung des EOS-Markers (end-of-sentence), die in diesem syntaktischen Kontext völlig ungrammatisch ist. Zum anderen aus der Aktivierung von Verben mit falscher Kongruenz. Beide Fehler werden über fortschreitende Epochen reduziert.
- Position des finiten Matrix-Verbs in SRS (in Bsp. 4b: *admitted*). Sowohl für SRS als auch für ORS war der Fehler auf dem Matrix-Verb sehr hoch. Die etwa gleiche Verteilung von Aktivierung auf Verben mit korrekter und mit falscher Kongruenz über mehrere Epochen weist auf unzureichend gelernte Fernabhängigkeiten hin. Die Hauptquelle des Fehlers in SRS liegt neben der zu geringen Aktivierung der Verben wiederum mit starkem Gewicht in der ungrammatischen Aktivierung des EOS-Markers. Diese Fehlerquellen änderten sich auch kaum über fortschreitende Epochen.
- Position des finiten Matrix-Verbs in ORS (in Bsp. 4a: *admitted*). Auch für die ORS waren die zu geringe Aktivierung der Verben sowie die ungrammatische Aktivierung des EOS-Markers Hauptquellen des Fehlers. Hinzu kommt mit starkem Gewicht die ungrammatische Aktivierung des Artikels (*the*).
- Multidimensionale Skalierung für die gemittelten Hidden-Layer-Vektoren für unterscheidbare Wörter: Der zentrale Befund aus der Analyse der internen Repräsentation betrifft das Relativpronomen (*that*). Dieses ist in der frühen Epoche im Repräsentationsraum noch mit größerer Nähe zu den transitiven Verben verortet und entfernt sich erst zu späterer Epoche von diesen.

Ein erster Aspekt der Erklärung der Simulationsergebnisse betrifft die Aktivierung des EOS-Markers auf der Position des finiten Relativsatz-Verbs in ORS: Dieser wird auf eine Fehlinterpretation des Relativpronomens aufgrund der mangelnden Distinktion von transitiven Verben zurückgeführt. Die Ähnlichkeit zwischen transitiven Verben und dem Relativpronomen beruhen auf einer weitgehenden Übereinstimmung ihrer Kookkurrenz im zugrundeliegenden Trainingskorpus: Zumindest im lokalen Kontext erscheint das Relativpronomen auf Positionen, auf denen auch Verben erscheinen können (Positionierung zwischen zwei Nominalphrasen). Damit wird das Netzwerk auf eine falsche Fährte für die Vorhersage geführt, denn ihm liegt ein unzutreffender syntaktischer Kontext für die Folgevorhersage zugrunde. Folglich ist nach der unzutreffenden Interpretation des Teilsatzes *The reporter that the senator ...*, in der *that* als Verb erfasst wird, tatsächlich ein Satzabschluss mit dem EOS-Markers möglich. Die Ursache für diese Fehlleistung ist also schlicht der unausgereifte Zustand des Netzwerks in der frühen Epoche bzw. die unzureichende Kategorisierung der sprachlichen Einheiten. Dies führt zu einer Wortvorhersage auf Grundlage eines falschen angenommenen syntaktischen Kontexts. Erst nach weiterem Training ist die interne Repräsentation genügend ausdifferenziert, damit der Unterschied zwischen Verben und dem Relativpronomen erfasst und diese Fehlinterpretation vermieden wird.

Ein zweiter Aspekt der Erklärung betrifft die Aktivierung des EOS-Markers auf der Position des finiten Matrix-Verbs in beiden Satztypen (sowie des Artikels in ORS): In beiden Fällen wird eine Erklärung herangezogen, die auf der Annahme eines lokal be-

grenzten Fensters der Verarbeitung beruht. Das bedeutet, dass nicht der gesamte Satz als syntaktischer Kontext für die Vorhersage des Folgeworts herangezogen wird, sondern nur ein lokaler, begrenzter Ausschnitt des Satzes. Die dadurch hervorgerufenen Aktivierungen falscher Units entsprechen dem als *false alarms* bezeichneten Anteil im globalen Fehler GPE (vgl. Kap. 3.2.1.1), die wirkenden Effekte werden hiernach als False-Alarms-Effekte bezeichnet. Im Falle der SRS werden die False-Alarms-Aktivierungen auf dem Matrix-Verb aus der lokal beschränkten Interpretation dieser Position als Abschluss eines einfachen transitiven Hauptsatzes erklärt (... *attacked the senator eos*). Im Falle der ORS vermuten die Autoren, dass die False-Alarms-Aktivierung des EOS-Markers auf der unzureichenden Unterscheidung zwischen transitiven und intransitiven Verben in Kombination mit der lokal beschränkten Interpretation der Matrix-Verb-Position als Ende eines intransitiven Hauptsatzes beruht (*the senator attacked eos*). Die False-Alarms-Aktivierung des Artikels hingegen entspricht der lokal beschränkten Interpretation des Teilsatzes als einfachem Hauptsatz, in dem auf das Verb ein Objekt mit vorausgehendem Artikel erscheinen muss (*the senator attacked the ...*).

**Fehler durch lokal konsistente Verarbeitung** Konieczny, Ruh und Müller (2009) identifizieren lokal konsistente False Alarms als Hauptquelle der Verarbeitungsschwierigkeiten in SRNs. Damit demonstrieren sie den Stellenwert der Verwendung von Details der Output-Aktivierung, um den ursächlichen Eigenschaften und Mechanismen der syntaktischen Verarbeitung in SRNs auf die Spur zu kommen. Die gefundenen Aktivierungs-Details müssen aus der Perspektive des Erklärungsansatzes von MacDonald und Christiansen (2002) unplausibel bleiben und stellen damit den Erklärungsansatz zumindest teilweise in Frage. Eine Analyse, die sich nur auf die Interpretation des globalen Fehlers beruft, beinhaltet also die Gefahr der Unzulänglichkeit und eventuell sogar der Fehlinterpretation. Allerdings bedeutet das nicht, dass Konieczny und Kollegen SRNs aufgrund ihrer kritischen Analyse als mögliches Modell menschlicher Satzverarbeitung ablehnen. Vielmehr bestätigen sie, dass verschiedene Aspekte des Netzwerkverhaltens grundsätzlich gute Ansätze bietet, empirisch beobachtbare Zusammenhänge zu erklären. Nur gilt es, die dabei wirkenden Mechanismen genauer zu identifizieren. Erfahrung bleibt dabei ein wichtiger Faktor, der Verarbeitungsunterschiede und interindividuelle Unterschiede erklären kann: Bezogen auf SRNs bestimmt die Erfahrung die Trennschärfe der Kategorienbildung in den internen Repräsentationen und damit die Präzision der Repräsentation des syntaktischen Kontexts und der darauf beruhenden Wortklassenvorhersagen (Konieczny & Müller, 2010).

Aus den Simulationsergebnissen schließen Konieczny, Ruh und Müller (2009) zusätzlich einen weiteren möglichen Faktor syntaktischer Komplexität: Interferenz-Effekte des lokalen Kontexts auf die Verarbeitung des globalen Kontexts und die fehlerhafte Antizipation der syntaktischen Struktur aufgrund lokal kohärenter (aber global inkohärenter) Teilsatzstrukturen. Auf einen solchen Zusammenhang verweisen auch Tabor et al. (1997) in einer Simulationsstudie sowie Tabor et al. (2004) in einer empirischen Studie, die Befunde zur psychologischen Realität solcher Interferenz-Effekte liefert. Konieczny (2005) untersucht auf dieser Grundlage, ob diese Interferenz-Effekte möglicherweise tatsächlich False-Alarms-Effekte sind – d.h. ob die Aktivierung falscher sprachlicher Folgeelemente aufgrund der Interferenz lokal kohärenter syntaktischer Kontexte psychologische Realität besitzt. In einer umfassenden Serie von Studien mit verschiedenen psycholinguistischen

Methoden wurde der Einfluss solcher Effekte lokaler Kohärenz auf die Sprachverarbeitung nachgewiesen (vgl. u.a. Konieczny, 2005; Konieczny & Müller, 2007; Müller, Hachmann & Konieczny, 2009; Konieczny, Müller, Hachmann et al., 2009; Weldle, Konieczny, Müller et al., 2010a; Konieczny & Müller, 2010; Konieczny, Weldle, Haser et al., 2010). Eine ausführliche Diskussion dieses psycholinguistischen Gegenstands mit Erläuterungen der Effekte und ihrer Bedeutung für Modelle der menschlichen Sprachverarbeitung sowie weiteren Erklärungsansätzen finden sich in Kapitel 5.2.

Fassen wir zusammen, welche weiteren Eigenschaften für SRNs festgestellt und als Erklärungsansätze für Eigenschaften der menschlichen Satzverarbeitung vorgeschlagen werden: Die Verarbeitung von Sätzen erfolgt nicht zwingend auf jeder Position in global kohärenter Weise, sondern wird parallel durch die kohärente Verarbeitung innerhalb eines begrenzten lokalen Fensters bestimmt. Dieses lokal beschränkte Verarbeitungsfenster entsteht in SRNs aus der Eigenschaft ihrer beschränkten Kapazität bei der Repräsentation des sequenziellen Kontexts sowie aus der starken Interferenz lokaler Informationen. Konieczny, Ruh und Müller (2009); Konieczny, Müller, Hachmann et al. (2009); Konieczny und Müller (2010) legen weitere Erklärungsmuster vor, die hierbei relevant werden könnten. Sie verweisen unter anderem darauf, dass SRNs in erster Linie Modelle des Spracherwerbs bzw. Lernmodelle sind. Dabei wurde in verschiedenen Studien darauf hingewiesen, dass im Verlauf des Erwerbs komplexere Strukturen wahrscheinlich auf Grundlage zuvor erworbener einfacherer Strukturen erworben werden (vgl. Elman, 1993; Rohde & Plaut, 1999). Die Fokussierung auf lokale Fenster bei der Verarbeitung sowie die daraus resultierenden Effekte lokaler syntaktischer Kohärenzen könnten deshalb ein Überbleibsel der zugrundeliegenden funktionalen Spracherwerbsmechanismen sein.

Als weiterer relevanter Aspekt wird die ressourcenfreie Parallelverarbeitung in SRNs hervorgehoben.<sup>37</sup> Dies könnte erklären, weshalb Antizipationen für multiple parallele Satzfolgen und sogar lokale Vorhersagen verarbeitet werden, was in rationalen Parsingmodellen als irrational und nicht sparsam abgelehnt werden müsste (vgl. dazu die Diskussion in Kap. 5.2.4). Dabei ist das Verhalten, auch Vorhersagen innerhalb des lokalen Kontexts zu treffen, möglicherweise sogar durchaus sinnvoll und kognitiv ökonomisch, da die lokal gegebenen Informationen einen großen Teil der strukturellen Bezüge aufklärt (vgl. Demberg & Keller, 2009). Hinsichtlich dieser in den meisten Fällen erfolgreichen Strategie der Nutzung lokaler Informationen für die Antizipation der Folgestruktur von Sätzen vermuten Konieczny, Müller, Hachmann et al. (2009), dass SRNs eine gute Implementierung des Good Enough Parsing-Ansatzes (Ferreira, Bailey & Ferraro, 2002; Ferreira & Patson, 2007) sind.

### 3.2.4. SRNs als probabilistische Parser

Abschließend wollen wir kurz besprechen, welcher Modellklasse SRNs mit Wortvorhersage-Aufgabe zuzuordnen sind. Sind SRNs Parser im klassischen Sinne? Diese Frage muss wohl mit Nein beantwortet werden, denn die Zielstruktur, die sich bei der Wortvorhersage

---

<sup>37</sup>Allerdings sollte hier hinterfragt werden, ob die Verarbeitung in SRNs wirklich als komplett ressourcenfrei bezeichnet werden kann. Denn schließlich führt eine höhere Anzahl an Alternativen zu einer geringeren Präzision der positionsbezogenen Aktivierungsverteilung, einem höheren Fehler und einer schlechteren Antizipation auf Folgepositionen (vgl. z.B. Weckerly & Elman, 1992).

entwickelt, ist keine konsistente syntaktische Gesamtstruktur des Satzes.<sup>38</sup> In diesem Zusammenhang spricht Steedman (1999) SRNs den Status von Parsern ab und klassifiziert sie tendenziell als Part-Of-Speech-Tagger. Sind SRNs Modelle des Satzverstehens? Auch diese Frage muss wohl eher mit einem Nein beantwortet werden, denn die Modelle leisten keine Konstruktion einer kohärenten Gesamtrepräsentation des Satzinhalts. Hinsichtlich ihrer Eigenschaften des Erwerbs syntaktischer Strukturen, der sensitiv ist für distributionale Merkmale des Inputs, und hinsichtlich ihrer probabilistischen Vorhersage möglicher Folgewortklassen auf Grundlage des syntaktischen Kontexts, können SRNs möglicherweise treffender als Implementierung probabilistischer Parser und statistischer Lernmodelle charakterisiert werden. Unter dieser Perspektive wurden in verschiedenen Forschungsrichtungen Beziehungen zu konnektionistischen Simulationen mit SRNs hergestellt, die im Folgenden kurz zusammengefasst werden.

Spracherwerb und Sprachverstehen erfordern den Umgang mit unsicherer Information, die Inferenzen verlangen, um eine sichere und vollständige Analyse des Inputs zu gewährleisten. Der Umgang mit unsicheren Informationen kann sehr gut mit probabilistischen Modellen erfasst werden, die genau diese Unsicherheiten in den Daten abbilden. Dabei gibt es viele verschiedene Ansätze zur Modellierung, wobei in der jüngeren Vergangenheit den Bayesschen Modellen verstärkt Aufmerksamkeit gewidmet wird. Der probabilistische Ansatz an sich ist dabei in vielerlei Hinsicht theorieneutral bzw. kompatibel zu sehr verschiedenen Ansätzen (etwa auch gegenüber nativistischen und empiristischen Positionen, vgl. Chater & Manning, 2006). Der Nutzen probabilistischer Modelle für Fragen der Psycholinguistik besteht darin, dass sie einen Zugang zur Struktur des evidenzbasierten Schlussfolgerns eröffnen und mit Algorithmen zur Gewichtung und Kombination von Evidenzen in den Daten einen Ansatz für die Auswahl und Entscheidung für eine bestimmte Interpretation der Daten bieten (Jurafsky, 2003). Konnektionistische Modelle haben einen starken Bezug zu probabilistischen Modellen, auch wenn die in ihnen realisierten probabilistischen Annahmen nicht explizit formuliert sind (Chater & Manning, 2006). Im Unterschied zu formal formulierten Modellen, die eher auf der funktional-abstrakten Ebene der Erklärung kognitiver Prozesse operieren, sind konnektionistische Modelle als Realisierungen probabilistischer Modelle zwischen algorithmischer und implementationaler Ebene zu verstehen. Sie sind aus der Sicht probabilistischer Theorien statistische bzw. konnektionistische Methoden, um Daten und Strukturen nach probabilistischen Prinzipien zu verarbeiten und zu erwerben. Sie explizieren die kognitiven Prozesse, die für das probabilistische Verhalten in der menschlichen Performanz verantwortlich sind (vgl. Hale, 2003, S. 103): „Explicit mechanisms of probabilistic processing have been pursued most intently by connectionists.“ Allerdings ist die genaue Beziehung zwischen probabilistischen und konnektionistischen Theorien alles andere als klar: „Deciding whether probabilistic models are merely a higher-level description of connectionist models, or whether the two are mutually exclusive alternatives, remains a key problem for future research.“ (Jurafsky, 2003, S. 36).

---

<sup>38</sup>Zumindest wird keine explizite Repräsentation der Gesamtstruktur aufgebaut. Dem steht die Interpretation der Trajektorien durch den Repräsentationsraum (u.a. Elman, 1991) entgegen. Hier wird die Gesamtstruktur dadurch berücksichtigt, dass sich der aktuelle Zustand innerhalb eines Satzes sowohl aus dem aktuellen Input als auch aus den vorausgegangenen Zuständen im Satz ergibt. Damit ist gewissermaßen im aktuellen Zustand implizit der gesamte Satz repräsentiert.

Hinsichtlich des Spracherwerbs besteht ein klarer Zusammenhang zwischen SRNs und probabilistischen Modellen: SRNs nutzen statistische Lerneigenschaften, um anhand von spärlichem, unvollständigem und mitunter verrauschtem Input Generalisierungen über die zugrundeliegenden strukturellen Zusammenhänge der Daten zu leisten. Elman (2003) stellt hierzu fest, dass es beim Sprachlernen nicht nur um das Lernen von statistischen Merkmalen der Sprache geht, sondern um statistisch gesteuertes Lernen (vgl. Kap. 3.1.3.2): Im Zentrum der Leistungsfähigkeit der Modelle steht nicht Memorierung, sondern Induktion. Und die Modelle bieten einen Ansatz dafür, wie diese Induktion geleistet wird. In den Worten von Elman (2003): „Induction may be the name of the game, but constraints are the rules that we play by.“ Im vorausgegangenen Unterkapitel 3.2.2 wurden Ansätze erwähnt, die eine Beziehung zwischen konnektionistischen Netzwerken und empirischen Sachverhalten zum statistischen Lernen herstellen und spezifische, aus konnektionistischen Simulationen abgeleitete Aussagen darüber überprüft haben (Wells et al., 2009; Gennari & MacDonald, 2008; Misyak et al., 2008).<sup>39</sup>

Auch hinsichtlich der syntaktischen Verarbeitung bestehen deutliche Zusammenhänge: Die Vorhersage-Aufgabe führt zu einem Verhalten der SRNs, das als probabilistische Wortklassenvorhersage charakterisiert werden kann. Für alle möglichen Wörter der sprachlichen Umwelt des SRNs wird eine probabilistisch gewichtete Aktivierung produziert. Die Gesamtaktivierung entspricht einem Wahrscheinlichkeitsvektor für mögliche Folgewörter – auf Grundlage des gegenwärtigen Wortes im Kontext der vorhergegangenen Wörter der Sequenz und mit Sensitivität für die bisherige Erfahrung mit der sprachlichen Umwelt und deren statistisch-strukturellen Merkmalen. Die Eigenschaften der Antizipation in SRNs beruhen auf der Interaktion der distributionalen Eigenschaften des Inputs einerseits und den Eigenschaften der Architektur andererseits, die auf ihren Lernalgorithmen, ihrer Repräsentationsstrukturen aufgrund vorheriger Erfahrungen und ihrer dynamisch-rekurrenten Verarbeitung beruhen. Auf den Zusammenhang zwischen den Eigenschaften von SRNs und der Antizipation als gewichtigem empirisch nachweisbarem Merkmal menschlicher Satzverarbeitung haben wir an früherer Stelle hingewiesen (vgl. Kapitel 3.2.1 und 3.1.2). Aktuell entstehen verschiedene weitere Ansätze abseits konnektionistischer Modellierung, in denen Antizipation auf Grundlage probabilistischer Informationsverarbeitung eine zentrale Rolle spielt. Surprisal und Entropy Reduction sind informationstheoretische Operationalisierungen der angenommenen Prozesse (vgl. Hale, 2001; Charniak, 2001; Hale, 2003; R. Levy & Andrew, 2006; R. Levy, 2008; Boston, Hale, Kliegl, Patil & Vasishth, 2008; Demberg & Keller, 2009; Hale, 2011). Für diese wird eine starke Korrelation für Wortlesezeiten innerhalb von Sätzen angenommen, wobei der angenommene kognitive Aufwand durch den Informationsgehalt eines Wortes bzw. durch seine Vorhersagbarkeit oder Unsicherheit im syntaktischen Kontext bestimmt wird. Der Surprisal-Wert als Maß der Komplexität eines Folgewortes entspricht also seiner Unerwartetheit nach dem vorhergehenden Satzausschnitt. Für die Ableitung solcher Werte zum Vergleich mit empirischen Daten kommen verschiedene Modellierungsmethoden in Frage – auch SRNs. S. L. Frank (2009) vergleicht Surprisal-Werte, die korpusbasiert mit PCFGs sowie mit SRNs generiert wurden, und evaluiert diese anhand empirischer Daten für Lesezeiten aus Eyetracking-Studien. Er kommt dabei zu dem Schluss, dass SRNs ein besseres psycholin-

---

<sup>39</sup>Weitere Hinweise auf statistisch geleiteten Spracherwerb und empirische Befunde zu Effekten statistischer Lernmechanismen beim Sprachlernen im Artificial Language Learning-Paradigma finden sich bei Saffran (2001, 2002, 2003).

guistisches Modell darstellen, das die Vergleichsdaten aufgrund seiner Eigenschaften und Beschränkungen akkurater abbilden kann. Mit seinen Ergebnissen unterstützt er zudem die Annahme probabilistischer Mechanismen wie Surprisal und Entropy Reduction als psychologisch valide Mechanismen der inkrementellen Revision der Satzinterpretation während der Verarbeitung (vgl. auch S. L. Frank, 2010).

Es bestehen klare Zusammenhänge zwischen SRNs und probabilistischen Theorien der Sprachverarbeitung. So können etwa erstere als Implementierung letzterer dienen. Im Unterschied zu formalen Modellen haben SRNs dabei den Vorteil, explizit die Prozesse und Mechanismen zu modellieren, die den probabilistischen Effekten zugrunde liegen. Sie konkretisieren implizite Beschränkungen, verdeutlichen interne Repräsentationen und Zusammenhänge und simulieren Constraint-Satisfaction-Prozesse.<sup>40</sup> Sie bieten damit nicht nur einen Anhaltspunkt für die Beantwortung der Frage, ob in unserer Sprachverarbeitung probabilistische Mechanismen zum Tragen kommen, sondern auch, wie sie zum Tragen kommen.

---

<sup>40</sup>Auch im größeren Kontext allgemeiner Multiple-Constraint-Ansätze stellen SRNs damit ein Werkzeug bzw. eine Methode dar, mit der spezifische Hypothesen generiert, theoretische Annahmen überprüft und Eigenschaften von Multiple-Constraint-Ansätzen entdeckt und spezifiziert werden können (vgl. MacDonald & Seidenberg, 2006).



## 4. Konnektionistische Modelle des Sprachverstehens

Im folgenden Kapitel beschäftigen wir uns mit der Frage, wie in konnektionistischen Modellen Bedeutung und Verstehen im Sinne des Aufbaus eines kohärenten Situationsmodells konzipiert werden. Wir haben im vorhergehenden Kapitel zwischen SRNs mit Vorhersageaufgabe und SRNs mit Assoziations- oder Komplettierungsaufgabe unterschieden. In ersteren wird die Bedeutung für jeweils einzelne Elemente (i.d.R. Wörter) verarbeitet, wobei diese nur indirekt als Wahrscheinlichkeitsverteilung des Output-Layers und als Verortung im Repräsentationsraum für die Hidden-Layer erfasst werden kann. Die Konstruktion von Bedeutung erfolgt hier nur flüchtig. Die Bedeutung eines gesamten Satzes kann nur über die Trajektorie durch den Repräsentationsraum nachvollzogen werden. Unklar bleibt auch, inwiefern eine Revision bereits aufgebauter Bedeutung erfolgen und repräsentiert werden kann. Die Modelle der zweiten Klasse hingegen modellieren Sprachverarbeitung im Sinne der Konstruktion einer holistischen Bedeutung von Sätzen. Zumeist werden dazu Wortsequenzen auf partielle oder vollständige statische Strukturen abgebildet, mit dem Ziel, jedem Input eine kontextabhängige Rolle im Constraint-Satisfaction-Prozess bei der Konstruktion der vollständigen Bedeutung zuzuweisen. Natürlich existieren auch andere Modelle, in deren Architektur keine SRNs einbezogen werden. Die meisten Modelle bauen jedoch auf SRNs auf und unser Fokus liegt auf diesen Ansätzen.

In diesem Kapitel werden wir nach einer allgemeinen Einführung in drei Abschnitten einige Modelle und Entwicklungslinien skizzieren und hinsichtlich ihrer Architektur, ihrer zugrundeliegenden semantischen Repräsentation und ihrer Kontrollmechanismen bei der Konstruktion komplexer Bedeutungszusammenhänge vergleichen.<sup>1</sup> Dabei müssen die in diesen Ansätzen verfolgten unterschiedlichen Ziele berücksichtigt werden. Wir besprechen das SG-Modell (McClelland et al., 1989; St. John, 1992), das CSCP-Modell (Rohde, 2002), das INSOMNet-Modell (Mayberry, 2003), das DSS-Modell (S. L. Frank et al., 2003b, 2009), das CIA-Net (Mayberry, Crocker & Knoeferle, 2005; Mayberry et al., 2009) sowie einige andere einfach erweiterte SRN-Modelle (Dominey & Boucher, 2005; Altmann & Kamide, 2009). In einem abschließenden Unterkapitel wird es darum gehen, aus übergeordneter Perspektive nachzuvollziehen, wie Bedeutung und Sprachverstehen in diesen Ansätzen konzeptualisiert sind und welche Perspektive dabei auf Satzsemantik gelegt wird. Wir werden den besprochenen Modellen nochmals in Kapitel 9 begegnen, in dem wir sie aus einem spezifischen Blickwinkel kritisch diskutieren.

---

<sup>1</sup>In diesem Kapitel werden ausschließlich rein konnektionistische Modelle im Sinne des eliminativen Konnektionismus besprochen, nicht aber hybride oder strukturiert konnektionistische Modelle.

## 4.1. Modelle und Entwicklungslinien

Die Bedeutung von Sätzen geht weit über das hinaus, was als Summe der Bedeutungen der einzelnen beteiligten Wörter erfasst werden kann. Neben den einzelnen lexikalischen Bedeutungen müssen deren Verweise und Beschränkungen aufeinander erfasst und aus den syntaktischen Informationen der relationale, kausale, oder raum-zeitliche Zusammenhang zwischen den benannten Elementen abgeleitet werden. Weiterhin ist der Einbezug kontextueller und pragmatischer Informationen sowie des allgemeinen Weltwissens erforderlich, um die tatsächliche Bedeutung eines Satzes zu erfassen. Welche Repräsentationsstruktur kann dieser reichhaltigen Konstruktion von Bedeutung Rechnung tragen? Und welche Repräsentationsstruktur wird dem gerecht, was wir Menschen letztendlich als Bedeutung eines Satzes kognitiv verstehen? In der Linguistik wurden verschiedene Modelle für Satzbedeutung und zur Formalisierung des semantischen Gehalts von Diskurs und Text vorgeschlagen. Gleichfalls wurden in der Psychologie, der Psycholinguistik und der Kognitionswissenschaft verschiedene Modelle des Sprachverstehens entwickelt, so etwa das Structure Building Framework (Gernsbacher, 1990), das Constructivist Framework (Graesser, Singer & Trabasso, 1994), das Event-Indexing-Modell (Zwaan & Radvansky, 1998), das Immersed Experiencer Framework (Zwaan, 2004) und, sicherlich am prominentesten, das Construction-Integration-Modell (Kintsch, 1998; Kintsch & van Dijk, 1978).

Im Großen und Ganzen ist das Feld der Theorien zur Satz- und Diskurs-Bedeutung von symbolischen Ansätzen der Repräsentation und Verarbeitung beherrscht. Allerdings wurden auch nicht-symbolische bzw. subsymbolische Realisierungen entwickelt. Neuere Ansätze in der Tradition der Embodied Cognition werden in Kapitel 8.2 besprochen, Modellierungen zu diesem Ansatz in Kapitel 9. In diesem Kapitel beschäftigen wir uns mit konnektionistischen Ansätzen sprachlicher Bedeutung, speziell der Bedeutung von Sätzen. Dabei ist auch für diese Modelle festzustellen, dass sehr unterschiedliche Repräsentationen zugrundegelegt und unterschiedliche Mechanismen und Transformationsstrukturen angenommen werden. Man kann also kaum von einer kohärenten konnektionistischen Theorie der Satzbedeutung im Sinne eines einheitlichen Ansatzes sprechen. Einige Funktionsweisen und der generelle Apparat konnektionistischer Verarbeitung und Repräsentation werden geteilt, und auch der grundlegende Gegenstand – der kognitive Prozess der Ableitung einer zum Satz korrespondierenden Bedeutung – ist derselbe. Für die spezifischen Eigenschaften und Funktionsweisen hingegen, die für das jeweilige Modells entscheidend sind, kommen sehr unterschiedliche Annahmen zur Geltung. Wir skizzieren kurz einige zentrale Unterschiede.

Mit den verschiedenen Ansätzen werden unterschiedliche Ziele verfolgt. Einige Modelle zielen primär auf die Simulation psycholinguistischer oder neuropsychologischer Daten aus verschiedenen Experimentalparadigmen ab. Andere Modelle versuchen, speziell Aspekte des Sprach- und Bedeutungserwerbs und der Abbildung von sprachlicher Form auf Bedeutung zu erfassen. Damit dienen sie mitunter auch der komputationalen Fundierung von Grammatiktheorien, etwa der Konstruktionsgrammatik. Des Weiteren können kleinskalierte Modelle von hochskalierten Modellen unterschieden werden, anhand derer gezeigt werden soll, dass konnektionistische Modelle einen ernstzunehmend großen Ausschnitt der Sprache verarbeiten können. Und nicht zuletzt zielen manche Modelle spe-

ziell darauf ab, den Grad semantischer Systematizität bei der Verarbeitung zu untersuchen, den konnektionistische Modelle leisten können. Wir können aus den Ansätzen innerhalb der unterschiedlichen theoretischen Bezugsrahmen Aussagen über die Konzeption von Bedeutung ableiten. Die unterschiedlichen Leistungen der Modelle müssen aber in erster Linie vor dem Hintergrund ihrer jeweiligen Ziele beurteilt werden.

Abhängig von den jeweiligen Zielen wird ein unterschiedlicher Ansatz für Bedeutung gewählt. Dabei ist zu unterscheiden, was in den Modellen einerseits tatsächlich als kognitives Substrat der Bedeutung erachtet und welcher Komponente des Modells die Realisierung und Aktivierung der Bedeutung zugewiesen wird, und was den Modellen andererseits über die Form der Input-Output-Strukturen für die Verarbeitung und den Erwerb der Repräsentationen zugrundegelegt wird. Allerdings muss klargestellt werden, dass mit der Form der Input-Output-Struktur bereits entscheidende Aussagen über den Modellierungsansatz und die Konzeption von Bedeutung getroffen werden. Denn durch sie wird bestimmt, welche Strukturen, Verteilungen und Zusammenhänge in der Umwelt des Modells zugegen sind. Durch Form der Input-Output-Struktur wird auch bestimmt, welche Aufgabe das Netzwerk erfüllen soll – simple Assoziation, Antizipation oder Autoassoziation mit Komplettierung vorgegebener Schemata. Über diese vorgegebenen Strukturen und Aufgaben wird stark beeinflusst, welche Eigenschaften und Zusammenhänge die Modelle in ihren internen Strukturen abbilden. Meist wird damit auch indirekt eine Aussage darüber getroffen, welche Begriffe von Bedeutung und Verstehen diesen Modellen zugrunde liegen. So zielen manche Ansätze darauf ab, ein Parsing syntaktischer Strukturen im Sinne einer Analyse syntaktischer Strukturen über hierarchische Abhängigkeiten oder thematische Rollen zu leisten. Andere versuchen, die Konstruktion von Bedeutungsmodellen zu simulieren, sei es im Sinne propositionaler Strukturen oder mentaler Situationsmodelle. Auch der Erwerb syntaktischer Strukturen als bedeutungshafte Einheiten im Sinne der Konstruktionsgrammatik (Form-Bedeutung-Abbildungen) wird verfolgt. Die Konstruktion erfolgt mitunter als Ausfüllen propositionaler Strukturen oder thematischer Rollen im Sinne der Besetzung freier Positionen in vorgegebenen Schemata. In anderen Ansätzen hingegen erfolgt keine strukturelle Vorgabe, die Abbildung sprachlicher Elemente richtet sich hier auf holistische situationale Strukturen. Je nachdem, welche Strukturen vorgegeben werden und welche Aufgaben bei der Bedeutungskonstruktion erfüllt werden sollen, unterscheiden sich die Ansätze auch in den Mechanismen oder Kontrollstrukturen, über die die korrekte Konstruktion der zur Aussage korrespondierenden Bedeutung geleitet wird. Diese Kontrollstrukturen spiegeln sich zumeist direkt in der Architektur der Netzwerke, in der Aufbereitung der Input-Output-Strukturen sowie im systematischen Aufbau und Ablauf des Trainings wider. So finden sich unter anderem Abfragemechanismen für Positionen in Schemata, Kontrollmechanismen über selbstorganisierte Karten, komplexe modulare Netzwerktopologien und gestaffelte modulare Trainingsregimens (d.h. verschiedene Komponenten werden getrennt und weitgehend unabhängig voneinander trainiert).

Wir versuchen in den folgenden Unterkapiteln, zentrale konnektionistische Ansätze zusammenzufassen und Bezüge zwischen diesen hinsichtlich grober Entwicklungslinien und gemeinsamer Konzepte herzustellen. Die Charakterisierung der Ansätze richtet sich nach folgenden Gesichtspunkten:

- Welche Grundidee und welche Ziele werden mit dem Ansatz verfolgt, und vor welchem Hintergrund sind die Leistungen des jeweiligen Modells zu bewerten?

- Welche Form wurde für die Input-Output-Strukturen gewählt, welche Aufgaben muss das Netzwerk erfüllen und wie erwirbt das Netzwerk sprachliches Wissen?
- Welche Architektur wurde für das Modell entwickelt, wie erfolgt darin die Kontrolle der Konstruktion der Gesamtbedeutung von Sätzen und welche Modularisierung bzw. welches Trainingsregimens wird dafür vorgegeben?
- Wie wird sprachliche Form auf Bedeutung abgebildet, welches Substrat der Aktivierungen wird als Kern der Bedeutung aufgefasst und in welcher Komponente der Architektur wird er generiert?

Um ansatzweise übergreifende Konzeptionen konnektionistischer Sprachverstehensmodelle zu erfassen, unterscheiden wir Modelle mit einem Abfragemechanismus, SOM-gestützte Modelle und Modelle, die eine direkte assoziative Abbildung herstellen. Ausführliche Reviews über eine größere Anzahl konnektionistischer Sprachverstehensmodelle wurden von Rohde (2002) und Mayberry (2003) erstellt. Eine kritische Besprechung aktueller Modelle (speziell Rohde, 2002; Mayberry, 2003; S. L. Frank, Koppen, Noordman & Vonk, 2003a; S. L. Frank et al., 2007) einschließlich semantiktheoretischer Hintergründe findet sich in der Magisterarbeit von Wolfer (2007).<sup>2</sup>

## 4.2. Bedeutungskonstruktion durch Abfragemechanismen

In diesem Abschnitt diskutieren wir das Sentence-Gestalt-Modell (SG-Modell, McClelland et al., 1989; St. John & McClelland, 1990; St. John, 1992; St. John & McClelland, 1992) und das Connectionist Sentence Comprehension and Production-Modell (CSCP-Modell, Rohde, 2002). Das CSCP-Modell ist eine elaborierte Fortsetzung des im SG-Modell verfolgten Ansatzes und teilt mehrere Eigenschaften mit diesem: (a) Eine propositionale Slot-Filler-Repräsentation, die dem Aufbau der semantischen Repräsentation zugrundegelegt wird, (b) die Idee der Kompression sukzessive dargebotener propositionaler Informationen in einem sogenannten Gestalt-Layer, in dem die zusammenhängende Gesamtbedeutung von Sätzen konstruiert wird und (c) die Verwendung eines Abfragemechanismus, mit dem die Konstruktion und Kompression der Gesamtsatzbedeutung kontrolliert wird.

### 4.2.1. Sentence/Message Gestalt

Mit dem SG-Modell (McClelland et al., 1989; St. John & McClelland, 1990; St. John, 1992) soll die Zuweisung von Konstituenten aus Sätzen bzw. Texten zu thematischen Rollen als Constraint-Satisfaction-Prozess simuliert werden. Dabei soll das Modell mehrere

---

<sup>2</sup>Wir werden uns bei der Beschreibung der Modelle in diesem Kapitel stark beschränken. Weitergehende technische Details wie zum Beispiel die Größen der Layer oder die Trainingsparameter sowie Besprechungen detaillierter Performanzanalysen und Experimente werden nicht behandelt. Diese Details sind unseres Erachtens für die Erfassung der grundlegenden Idee und der Konzeptionierung des Verstehensprozesses irrelevant und damit für die weiteren Diskussionen dieser Arbeit nicht von Bedeutung.

Probleme konventioneller symbolisch-propositionaler Ansätze lösen<sup>3</sup> und dafür eine Reihe von Leistungen erfüllen: (a) Die Abrufbarkeit explizit dargebotener Propositionen, (b) die Auflösung pronominaler Referenzen, (c) die Disambiguierung ambiger und Konkretisierung vager Konstituenten, (d) die Herstellung von Kohärenz für den erfassten Text und das Treffen von Vorhersagen für erwartete Konstituenten durch Inferenzen, (e) die Revision bereits konstruierter Interpretationen aufgrund neu hinzukommender Informationen, (f) die Generalisierung von Wissen über Konstituenten aus spezifischen Kontexten auf neue Texte (alle bezogen auf das spätere Modell von St. John, 1992).

Im Modell wird das Repräsentationsprinzip rollenspezifischer propositionaler Argumentstrukturen umgesetzt, wobei die Argumentpositionen der Propositionen über verschiedene Abschnitte des Output-Layers (Slots) dargeboten werden. Als mögliche Argumente einer Proposition wurden Agent, Handlung, Patient, Objekt, Empfänger, Instrument, Art und Weise, Verortung und Handlungsbegleiter aufgenommen. Die Grundidee der Verwendung solcher Slot-Filler-Repräsentationen, mit denen propositionale Strukturen repräsentiert werden, gehen auf den Ansatz von Hinton (1986) zurück. Durch die Verknüpfung verschiedener Input- und Output-Layer, die semantische Konzepte im Sinne propositionaler Elemente wie zum Beispiel thematische Rollen oder Relationen repräsentieren, kann semantisches Wissen über die Eigenschaften und Zusammenhänge dieser semantischen Konzepte durch die Anpassung der Gewichtsstrukturen in verteilten internen Repräsentationen gespeichert werden. Anhand rollenspezifischer propositionaler Tripel (z.B. A IST-EIN B, A BESITZT-EIGENSCHAFT B, A HAT-FÄHIGKEIT B) können mit konnektionistischen Netzwerken semantische Netze (vgl. Collins & Quillian, 1969; Collins & Loftus, 1975) simuliert werden (Hinton, 1986; Rumelhart, 1990; McClelland & Patterson, 2002; Rogers & McClelland, 2003), die ihre Eigenschaften erfahrungsbasiert erwerben. Die Modelle entwickeln dabei hierarchische Strukturen mit Akkumulation und Vererbung propositionaler Merkmale und entwickeln inferenzielle Eigenschaften. Das SG-Modell geht einen Schritt weiter: Auf Grundlage sukzessive präsentierter Konstituenten werden in einem speziell trainierten Hidden-Layer – dem Sentence Gestalt-Layer (SG-Layer) – komplette Propositionen abgebildet (bezogen auf das Modell von McClelland et al., 1989). Im erweiterten Modell von St. John (1992) werden mehrere sukzessive präsentierte Propositionen im SG-Layer auf vernetzte, aufeinander bezogene Konjunktionen dieser Propositionen abgebildet. In den zuvor genannten Modellen semantischer Netze werden nur einzelne Propositionen erfasst, deren Konstituenten lokalistisch auf den Input- und Output-Layern repräsentiert werden, wobei die internen Repräsentationen die semantischen Beziehungen und Regelmäßigkeiten zwischen den Konstituenten abbilden. Im SG-Modell werden ganze Propositionen und mehrere aufeinander bezogene Propositionen erfasst, wobei die internen Repräsentationen sowohl die Konstituenten als auch deren Beziehungen und die aktuell intendierten Beziehungen zwischen den Propositionen abbilden müssen.

Die Architektur des SG-Modells besteht aus zwei Komponenten, einer Input- und einer Kontroll-Komponente, die über die gemeinsame Schnittstelle des Sentence Gestalt-Layers interagieren (vgl. Abb. 4.1). In der einfacheren Version des Modells (McClelland et al.,

---

<sup>3</sup>Als Hauptprobleme konventioneller Ansätze werden die Ambiguität, Vagheit, Gradualität, Kontextabhängigkeit und die Erfordernis von Inferenzen zu impliziten Informationen diskutiert, die sich bei der Verarbeitung natürlicher Sprache ergeben, und die es praktisch unmöglich machen, die Bedeutung eines Satzes alleine aus struktursensitiven Regeln abzuleiten.

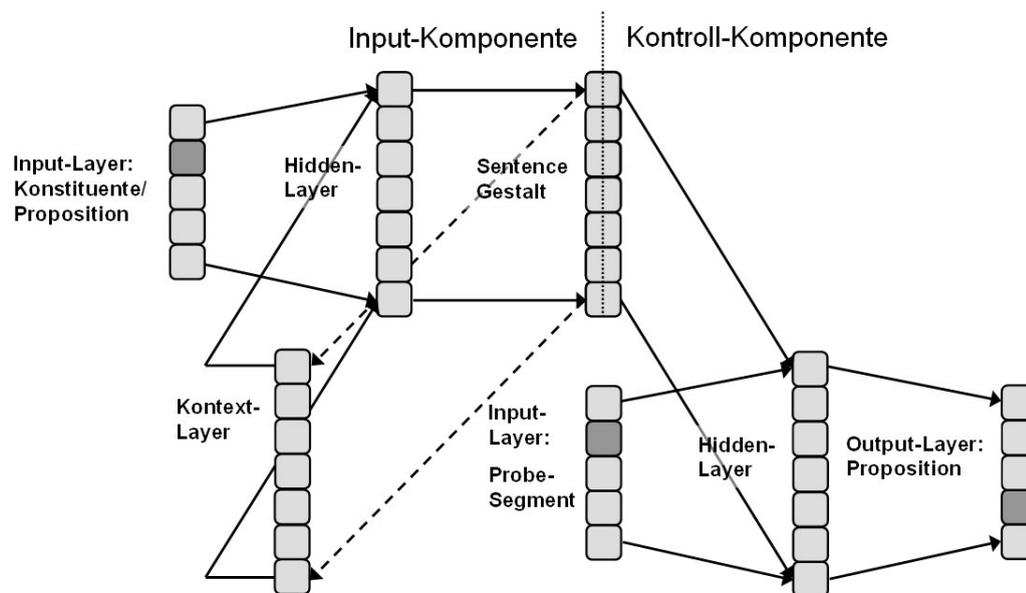


Abbildung 4.1.: Architektur des SG-Modells mit einem SRN als Input-Komponente und einem FFN als Kontroll-Komponente. Die Komponenten interagieren über den Sentence Gestalt-Layer, in dem die Satzbedeutung je nach Modell-Version als Proposition oder als vernetzte Konjunktion von Propositionen konstruiert wird.

1989) fließt Aktivierung der dargebotenen Konstituenten von der Input-Komponente durch das Netzwerk zum SG-Layer, wobei durch den rekurrenten Layer (im Sinne des SRNs nach Jordan, 1986) das jeweils vorhergegangene Aktivierungsmuster des SG-Layers mit eingespeist wird, um zu den neu hinzukommenden Konstituenten den Kontext des bisherigen Satzes zu integrieren. Durch den Aktivierungsprozess entwickelt sich ein spezifisches, nicht von außen determiniertes Aktivierungsmuster auf dem SG-Layer, das den gesamten Satz repräsentiert. Nachdem der ganze Satz eingespeist wurde, folgt der zweite Schritt: Für jede Konstituente des Satzes wird am Input-Layer der Kontroll-Komponente eine Frage zur thematischen Rolle oder zum Argument dieser Konstituente gestellt (d.h. die entsprechende Unit aktiviert). Zusammen mit der Satzrepräsentation auf dem SG-Layer, aus dem alle relevanten Informationen bezogen werden, muss auf dem Output-Layer die Frage beantwortet werden. Die Anpassung der Gewichte erfolgt über das gesamte Netzwerk durch den Backpropagation-Algorithmus. Dadurch wird die Repräsentation im SG-Layer optimiert und eine interne Struktur entwickelt, die eine Abbildung der Konstituenten auf die propositionale Struktur des Gesamtsatzes leisten kann. Das Modell von St. John (1992) unterscheidet sich insofern, als dass der Input aus mehreren sukzessive präsentierten vollständigen Propositionen besteht, und in der Kontroll-Komponente über unvollständige Propositionen die im Text beinhalteten Propositionen vollständig reproduziert werden müssen. Die Repräsentation im SG-Layer muss also (im Sinne einer komplexen Situation) mehrere Propositionen und deren Zusammenhänge abbilden.

Die Konstruktion von Satz- und Textbedeutung erfolgt im SG-Layer durch Constraint-Satisfaction-Mechanismen aus der Interaktion zwischen sequenziell eingelesenen Konstituenten bzw. Propositionen und der Vervollständigungsaufgabe für Ergänzungsfragen. Diese Form der Kontrolle der Bedeutungskonstruktion über den Abfragemechanismus weist

einen großen Vorteil auf: Es werden keine expliziten Vorgaben gemacht, welche Struktur die Satzbedeutung annehmen muss und wie die Vernetzung verschiedener Teilaspekte realisiert sein soll. Die Abbildung von Sprache auf spezifische Bedeutungsstrukturen wird damit durch die Abbildung von Sprache auf bestimmte Aspekte einer Aufgabe ersetzt. McClelland et al. (1989) beschreiben dies als Lernprozess einer impliziten Prüfung des Lernenden, die in etwa der Frage entspricht: ‚Sind die internen Repräsentationen und Verstehensmechanismen gut genug, dass die beschriebenen Ereignisse korrekt erfasst und folglich alle erforderlichen Fragen beantworten werden können? Die dabei intern aufgebaute verteilte Repräsentation der komplexen Bedeutungsstruktur muss dabei nicht direkt nachvollziehbar sein. Die Tatsache, dass sie das korrekte Antwortverhalten produziert, genügt als Beleg dafür, dass die Repräsentation vollständig ist und alle Zusammenhänge der beschriebenen Situation erfasst. Eine Analyse der Performanz zeigte gute Leistungen für mehrere der oben beschriebenen angestrebten Leistungen. Allerdings zeigten sich auch einige Schwachstellen des Modells sowohl hinsichtlich seiner Leistungsfähigkeit als auch seiner Leistungsmöglichkeiten: So war die Generalisierung auf neue Ereignisse nur sehr begrenzt erfolgreich. Zudem sind die Repräsentationsmöglichkeiten durch die gewählte Umsetzung der propositionalen Slot-Filler-Struktur deutlich beschränkt und erlauben unter anderem nur begrenzt Möglichkeiten zur Abbildung rekursiver Strukturen.

### 4.2.2. Connectionist Sentence Comprehension and Production

Auf den Arbeiten zum SG-Modell aufbauend stellt Rohde (2002) ein extensives, auf SRNs basierendes Modell des Sprachverstehens und der bedeutungsbasierten Sprachproduktion vor. Das Connectionist-Sentence-Comprehension-and-Production-Modell (CSCP-Modell) zielt in erster Linie darauf ab, ein großskaliertes konnektionistisches Modell der Sprachverarbeitung zu entwickeln, mit dem die Simulation empirischer Daten für ein breites Spektrum psycholinguistischer Phänomene innerhalb eines Modells geleistet werden kann, und mit dem die Skalierbarkeit konnektionistischer Ansätze und ihre Anwendbarkeit auf einen ernstzunehmenden Sprachausschnitt mit umfangreichem Lexikon und realistisch komplexer Grammatik belegt wird. Eine der Kernannahmen des Modells ist, dass Verstehen und Produktion von Sprache hinsichtlich der verstandenen oder produzierten Botschaft auf die gleichen Ressourcen zurückgreifen. Bei der inkrementellen Verarbeitung von Sätzen leistet das Modell einen Verstehensprozess, bei dem eine Message<sup>4</sup> generiert wird. Bei der Produktion leistet das Modell die Reproduktion einer syntaktischen Struktur, die zu einer vorgegebenen Message korrespondiert. Neben der Ausweitung auf die Sprachproduktion sollen mit dem Modell die Beschränkungen der verschiedenen Versionen des SG-Modells hinsichtlich der Komplexität der Bedeutungsstruktur und der abbildbaren syntaktischen Struktur überwunden werden.

Das CSCP-Modell besteht aus zwei distinkten Komponenten (vgl. Abb. 4.2): Das Message Encoder & Decoder-System (MED-System) und das Comprehension, Prediction & Production-System (CPP-System). Diese beiden Komponenten greifen auf einen gemeinsamen Pool von Informationen zu – die Message-Gestalt. Diese wird im zwischen

---

<sup>4</sup>Wir behalten im Folgenden den Begriff ‚Message‘ bei. Die Message entspricht in etwa der übertragenen Botschaft, dem semantischen Gehalt eines Satzes oder der strukturierten Bedeutung, die durch eine Äußerung übermittelt wird.

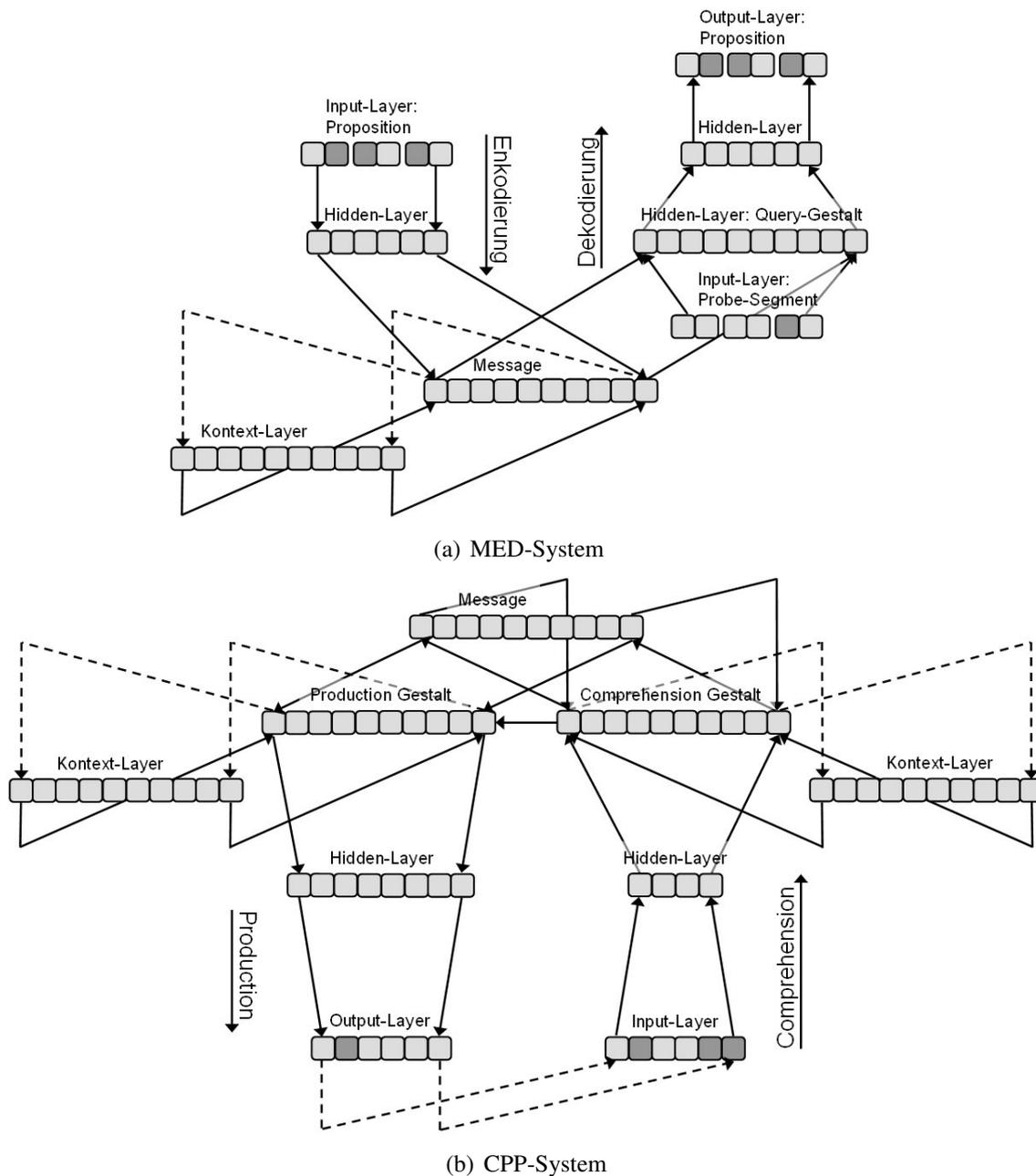


Abbildung 4.2.: Die Architektur des CSCP-Modells besteht aus zwei Komponenten: Oben wird das MED-System abgebildet, in dem anhand eines Abfragemechanismus die Message-Gestalt konstruiert wird. Unten wird das CPP-System abgebildet, das die Rezeption von Sprache, ihre Abbildung auf die Bedeutungsstrukturen und ihre Produktion modelliert. Die beiden Komponenten integrieren ihre Informationsstrukturen im zwischen-geschalteten Message-Layer, der als Schnittstelle zwischen sprachlicher Form und verteilt-propositionaler Bedeutung dient.

die Komponenten geschalteten Message-Gestalt-Layer aktualisiert, der die Schnittstelle zwischen sprachlicher Form und propositionaler Bedeutung bildet. Das Training der beiden Komponenten erfolgt getrennt: Zuerst wird das MED-System trainiert, um die komprimierten propositionalen Repräsentationen der komplexen Situationen zu generieren. Im zweiten Schritt wird das CPP-System mit sprachlichem Input trainiert, der zum einen auf sprachlichen Output (im Sinne der Wortvorhersage-Aufgabe) und zum anderen auf zum Satz korrespondierende komprimierte propositionale Repräsentationen abgebildet wird. Sprache und Bedeutung werden also getrennt bzw. unabhängig voneinander erworben und aufeinander abgebildet.

Bevor wir die beiden Hauptkomponenten noch einmal etwas näher beleuchten, besprechen wir kurz die Enkodierung der propositionalen Stimuli. Propositionen werden über Slots als relationale Tripel enkodiert. Diese Tripel bilden die Relationen zwischen Objekten, Handlungen, Zuständen, Eigenschaften, thematischen Rollen sowie einigen syntaktischen Markern kompositionell in den Slots ab: Argument A – Relation A/B – Argument B. Die semantische Enkodierung des Satzes *A cop saw that the young girl was bitten by that mean dog* etwa wird über sechs Propositionen enkodiert (vgl. Rohde, 2002, S. 120): Die erste Proposition P1 bezieht die Sinneswahrnehmung/-handlung *see* als Argument der Relation *Erlebender* auf das Argument *cop*. Proposition P1 besagt, dass ein Polizist irgendetwas sieht. Dabei werden auch weitere spezifizierende Merkmale der Argumente enkodiert, etwa dass der Akt des Sehens in der Vergangenheit stattfand, oder dass der Polizist indefinit ist und über die Eigenschaften menschlich, groß, alt, böse und männlich verfügt. Problematisch ist die Aufnahme syntaktischer Informationen als Bestandteil propositionaler Relationen, die syntaktische Aspekte des korrespondierenden Satzes abdecken: Die zweite Proposition P2 spezifiziert für Relation A/B mit *sc that* die syntaktische Relation der Argumente und weist Argument B die Rolle des Komplementsatzes (*sc*) für die in Argument A beschriebene Handlung zu. Mit diesem Vorgehen wird klar die Unterscheidung zwischen semantischer Bedeutung und syntaktischer Form verletzt. Dies ist vermutlich vor allem der Produktionskomponente geschuldet, in der ausschließlich aus der Message ein korrespondierender Satz generiert werden soll. Es dürfte ersichtlich geworden sein, dass diese Enkodierungsform sehr mächtig ist und sehr viel Information explizit ausweist. Die Propositionen wurden in Übereinstimmung zu den Trainingssätzen anhand einer probabilistischen kontextfreien Grammatik (PCFG) mit dem Simple Language Generator (SLG, Rohde, 1999b) generiert. Die PCFG modelliert die grammatisch und semantisch an der englischen Sprache orientierte Kunstsprache „Penglish“ (für pseudo- oder proto-Englisch, Rohde, 2002, S. 103). Diese Sprache wurde speziell dafür entworfen, die zu simulierenden psycholinguistischen Phänomene für die Modellierung zugänglich zu machen und die dafür benötigten Aspekte des Englischen zu erfassen. Hierfür wurden die entsprechenden lexikalischen und syntaktischen Strukturen implementiert und ihre Frequenzen statistisch anhand einer Korpus-Analyse der Penn Treebank (M. P. Marcus, Santorini & Marcinkiewicz, 1993) bestimmt.

Wie das Modell von St. John (1992) baut das CSCP-Modell seine Bedeutungsrepräsentation über einen Abfragemechanismus auf. Zu jedem Zeitschritt liegt eine Proposition an. In folgenden Zeitschritten können weitere Propositionen dargeboten werden, d.h. die Anzahl der Propositionen ist innerhalb gewisser Grenzen beliebig. Durch den rekurrenten Message-Layer werden die Propositionen in eine komprimierte und integrierte Form

überführt, die die Gesamtbedeutung der zusammengehörenden Propositionen repräsentiert (linke Hälfte des MED-Systems). Die Konstruktion dieser Repräsentation wird über den Abfragemechanismus im Dekodierungssystem geleitet (rechte Hälfte des MED-Systems): Als Probe-Segment liegt dazu eine unvollständige Proposition an, die am Output-Layer vollständig abgefragt wird. Die Vervollständigung muss durch die Informationen aus dem Message-Layer geleistet werden. Für jede Proposition wird jede Position des Tripels abgefragt. Bei jeder Folge-Proposition werden zu jedem Zeitschritt die vorangegangenen Propositionen auf diese Weise abgefragt, um sicherzustellen, dass im Message-Layer nicht nur die aktuelle Proposition enkodiert wurde, sondern die ganze zusammenhängende Folge von Propositionen, aus denen eine Message zusammengesetzt ist. Die Entwicklung der internen Repräsentationen auf dem Message-Layer erfolgt in der Interaktion von Enkodierungs- und Dekodierungssystem durch die Propagierung des Fehlers vom Output-Layer des Dekodierungssystems durch das gesamte MED-System bis zum Input-Layer des Enkodierungssystems.

Die Sprachverarbeitung für die penguinischen Sätze erfolgt im CPP-System: Auf dem Input-Layer werden verteilt pseudo-phonologisch enkodierte Wortformen in Silben-Slots als dreisilbige Wörter dargeboten; Wörter, die hinsichtlich der Silbenanzahl nicht passen, wurden entsprechend angeglichen. Diese Darbietungsform soll es dem System ermöglichen, auch phonologische und morphologische Informationen wie etwa Beugungen und Suffixe nutzen zu können. Die Sätze werden sukzessive Wort für Wort dargeboten und auf zwei Target-Vektoren abgebildet: Die erste Abbildung erfolgt über den Comprehension Gestalt-Layer auf die statische Repräsentation des Message-Layers, der die zum Satz korrespondierende Bedeutungsrepräsentation vorgibt. Die zweite Abbildung erfolgt über den Production Gestalt-Layer auf die lokalistisch dargebotene Repräsentation des Folgeworts im Sinne der klassischen Wortvorhersage-Aufgabe auf dem Output-Layer des Produktionssystems (linke Hälfte des CPP-Systems). Die Aufgabe des Comprehension Gestalt-Layers und des Production Gestalt-Layers ist die Vermittlung zwischen den inkrementellen sprachlichen und den statischen semantischen Informationen.

Für die Verarbeitung in der Test-Phase werden die beiden Komponenten des Modells gekoppelt. So kann die für einen Satz produzierte Verstehensleistung einerseits hinsichtlich der Wortvorhersageleistung und andererseits hinsichtlich der Verstehensleistung in der Decoder-Komponente des MED-Systems untersucht werden. Die Performanz wurde zudem über das speziell für das CSCP-Modell entwickelte Maß der Simulated Reading Time (SRT) evaluiert, die auch die Veränderung des Message-Layers pro Wort als Schätzwert des kognitiven Aufwands der Verarbeitung miteinbezieht.<sup>5</sup> Die Entwicklung solcher Verbindungshypothesen, über die der Zusammenhang zwischen der Performanz des Modells und den simulierten psychologischen Eigenschaften hergestellt wird, ist bedeutsam

---

<sup>5</sup>Die Simulated Reading Time (SRT, Rohde, 2002) ist ein komplexeres Performanz-Maß, das die Performanz verschiedener Verarbeitungskomponenten verrechnet. Die SRT berücksichtigt für die Schätzung der Lesezeiten in empirischen Experimenten sowohl Effekte der morphosyntaktischen als auch der semantischen Komponente. Für die Wortvorhersage als morphosyntaktischem Anteil wird die Divergenz (negativer Logarithmus der Aktivierung der korrekten Output-Unit) getrennt für Wortstämme und Affixe berechnet. Für den semantischen Anteil wird der RMSE der Differenz des Message-Layers sowie die durchschnittliche Aktivierung des Message-Layers (als Maß für die aktiv gehaltene Informationsmenge) berechnet. Alle vier Komponenten werden auf einen Mittelwert von 1.0 skaliert und anschließend prozentual gewichtet.

für die konnektionistische Psycholinguistik (vgl. Kap. 3.2.1.1) und erfordert eine gute Begründung, da über sie die empirische Validierung des Modells erfolgt. Das Modell bildet viele interessante Eigenschaften der Satzverarbeitung und des Verstehens ab und zeigt eine sehr gute allgemeine Performanz. Für die Untersuchung wurde eine große Anzahl psycholinguistischer Phänomene simuliert und ausführlich in jeweils eigenen Kapiteln besprochen.<sup>6</sup> Unter den simulierten Phänomenen befinden sich verschiedene lokale Ambiguitäten (u.a. Hauptverb/reduzierter Relativsatz, Nominalphrase/Komplementsatz), Anbindungsambiguitäten (Adverb-Anbindung, Präpositionalphrasen-Anbindung), lexikalische Ambiguitäten, Komplexitätsphänomene (verschiedene Relativsatz-Strukturen) sowie Produktionsfehler und -einflüsse (Kongruenzfehler, syntaktisches Priming). Auch wenn dabei viele Befunde aus empirischen Studien repliziert werden konnten, ist die generelle Leistung des Modells und das Spektrum simulierter Effekte beeindruckend und bietet Anregungen zu empirischen Folgeuntersuchungen.

### 4.3. SOM-gestützte Strukturabbildung: INSOMNet

In diesem Abschnitt besprechen wir das Modell INSOMNet (Mayberry, 2003; Mayberry & Miikkulainen, 2003) – das Incremental Nonmonotonic Self-Organization of Meaning Network. Im Zentrum steht dabei vor allem die Verwendung von Self-Organizing Maps (SOMs, Kohonen, 1990) und verschiedenen Weiterentwicklungen davon (v.a. dem SARDNet, Mayberry & Miikkulainen, 1999), mit denen die Transformation sprachlicher Sequenzen in ein vektorisiertes Format des komplexen semantischen Repräsentationsformalismus Minimal Recursion Semantics (MRS, Copestake, Flickinger, Pollard & Sag, 2005) kontrolliert wird. Wie im CSCP-Modell (Rohde, 2002) soll auch mit INSOMNet die Skalierbarkeit konnektionistischer Modelle auf eine ernstzunehmende Grammatik gezeigt werden, mit dem Ziel natürlichsprachliche Korpora zu erfassen. Im Unterschied zu Rohdes Modell, das eine auf ihre Inhalte abprüfbar implizite Bedeutungsstruktur aufbaut, versteht Mayberry sein Modell aber eher als Parser, das die bei der Satzverarbeitung konstruierten diskreten Strukturen explizit zugänglich macht. Zugleich soll INSOMNet Performanzaspekte der Sprachverarbeitung abbilden und dabei distinktive Eigenschaften konnektionistischer Modelle aufrechterhalten.

Die Architektur des INSOMNet besteht aus drei Komponenten (vgl. Abb. 4.3): (a) dem Sequence Processor, der die inkrementelle Verarbeitung des sprachlichen Inputs leistet, (b) dem Frame Encoder/Decoder, der die Konstruktion der zum jeweiligen Satz korrespondierenden semantischen Repräsentation leistet, und (c) dem Frame Selector, der die Auswahl der zu aktivierenden Komponenten des Frame Encoders unterstützt. Der Sequence Processor (Abb. 4.3(a)) besteht aus einem SRN, das den sprachlichen Input sequenziell Wort für Wort in diskreten Zeitschritten verarbeitet. Über einen rekurrenten Layer aktiviert der sequenzielle Input die Frame Map des Frame Encoder/Decoder-Systems. Die Verarbeitung des sequenziell dargebotenen Inputs wird unterstützt durch ein SARDNet (Sequential Activation Retention and Decay Network, James & Miikkulainen, 1995; Mayberry & Miikkulainen, 1999). SARDNets entsprechen Self-Organizing Maps (SOMs, Kohonen,

---

<sup>6</sup>Hier wird wiederum deutlich, dass das Modell in erster Linie als Modellierungswerkzeug für psycholinguistische Daten und zur Simulation psycholinguistischer Experimente entwickelt wurde.

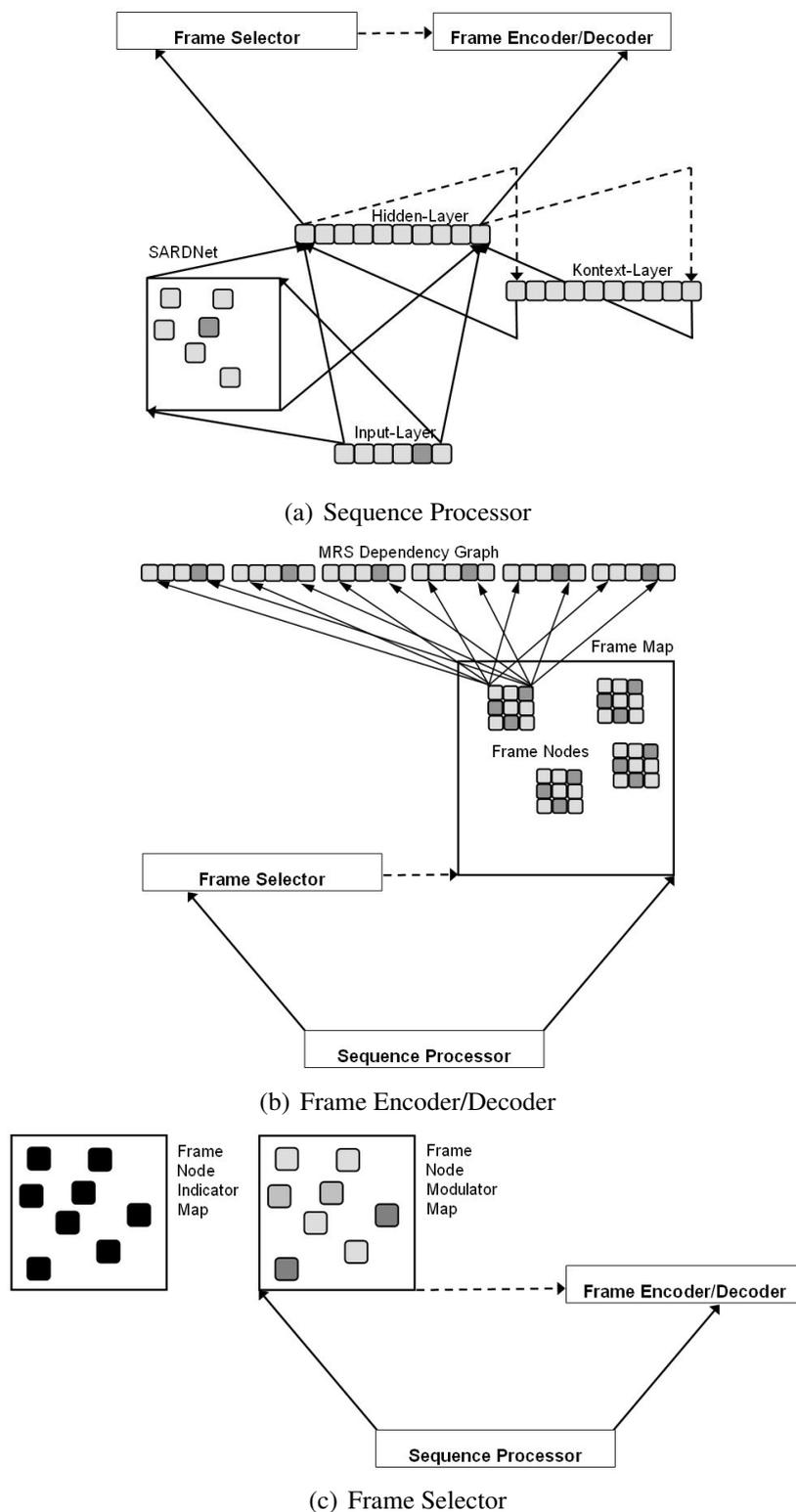


Abbildung 4.3.: Die Architektur des INSOMNet besteht aus drei Komponenten: dem Sequence Processor, dem Frame Selector und dem Frame Encoder/Decoder (modifizierte Abbildung nach Mayberry, 2003, S. 48ff).

1990)<sup>7</sup> für die Klassifikation sequenzieller Strukturen, die hinsichtlich ihrer sequenziellen Ähnlichkeit in topographischen Feldern klassifiziert werden. Das SARDNet speist Informationen über identifizierte Input-Muster in den rekurrenten Hidden-Layer ein und hilft dadurch, Fernabhängigkeiten zwischen Wörtern der Sequenz besser zu erfassen.

Bevor wir die beiden weiteren Komponenten besprechen, die die Konstruktion der Satzbedeutung konstituieren, beschreiben wir kurz den Semantikformalismus, auf dem die abgebildete Satzbedeutung begründet ist: Für die Konstruktion der semantischen Repräsentationen wird der Formalismus Minimal Recursion Semantics (MRS, Copestake et al., 2005) verwendet, ein Format, das die Abbildung komplexer hierarchischer Strukturen in einer flachen vektorisierbaren Repräsentation ermöglicht. Ähnlich wie in einer propositionalen Slot-Filler-Repräsentation umfasst die für INSOMNet gewählte Umsetzung der MRS mehrere distinktive Slots für die Komponenten, aus denen ein MRS-Frame jeweils zusammengesetzt ist. Ein MRS-Frame umfasst folgende Komponenten: (a) ein sogenanntes Handle, das etwa einer Adressierung des dargestellten Konzepts entspricht, (b) ein Konzept bzw. Wort oder Bedeutungsinhalt, (c) die Prädikation, die die Relation beschreibt, die durch das Konzept erfüllt wird, (d) eine Subkategorisierungs-Spezifikation, die die vom Konzept gebundenen Argumente beschreibt, und (e) eine mehrere Slots umfassende Liste der gebundenen Argumente.<sup>8</sup> Dadurch, dass die Handles in verschiedenen Frames als Argumente gebunden werden können, werden trotz der flachen Repräsentation komplexe hierarchische Strukturen und Zusammenhänge ermöglicht: Der Bezug auf andere Handles dient als Skopus, der die Abhängigkeiten zwischen verschiedenen Konzepten abbildet.

Das Frame Encoder/Decoder-System (Abb. 4.3(b)) bildet den sequenziellen sprachlichen Input auf die Gesamtbedeutung ab. Dies erfolgt in der Frame Map, die als zweidimensionale selbstorganisierende Karte von sogenannten Frame Nodes organisiert ist. Diese sind wiederum selbst zusammengehörende Verbände von Units. Die Frame Nodes repräsentieren die Bestandteile der Bedeutung des Satzes. Durch ihre Aktivierung durch den Sequence Processor entwickelt sich ein Aktivierungsmuster, das in der Dekodierung auf den MRS Dependency Graph abgebildet wird. Im MRS Dependency Graph wird jeweils der MRS-Frame dargeboten, der mit dem aktuellen Wort des sequenziellen Inputs korrespondiert. Welcher Frame Node jeweils welchen MRS-Frame bindet, wird nicht vorgegeben – die Zuweisung dieser Bindung erfolgt ähnlichkeitsbasiert durch die selbstorganisierte Frame

<sup>7</sup>Self-Organizing Maps (Kohonen, 1990) sind zweidimensional organisierte neuronale Layer, die eine selbstorganisierte Klassifizierung für Input-Muster leisten. Dies wird durch einen kompetitiven Lernalgorithmus bewirkt: Über exzitatorische und inhibitorische Gewichte zwischen den Units des Layers wird eine topographische Interpretation der Ähnlichkeitsstruktur des Inputs abgeleitet.

<sup>8</sup>Dies soll an einem kurzen Beispiel verdeutlicht werden (vgl. Mayberry, 2003, S. 42ff). Das Wort *hit* im Satz *The boy hit the girl with the doll* erhält damit einen Frame, der die folgenden Informationen enthält (vgl. Mayberry, 2003, S. 42ff): *h1* (Handle, Adresse der Repräsentation), *hit* (Wortform), *arg13\_rel* (transitives Verb), *A0A1A3DMEV* (Liste der Argumente, die durch das Wort gebunden werden). Die folgenden Positionen des Frames verweisen auf benannte Objekte oder MRS-Frames anderer Wörter, die als Argumente gebunden werden, optionale Argumente der Argumentliste können entfallen. Im Beispiel sind drei Argumente gebunden: *A1 x0* (Agens ist der Frame *x0* = boy), *A3 x1* (Patiens ist der Frame *x1* = girl), und *EV e0* (Handlungsart ist *e0* = Ereignis). Der gegebene Satz ist global ambig, denn er kann gleichermaßen bedeuten, dass die Puppe das Instrument der Handlung ist (d.h. mit der Puppe wird geschlagen), oder dass die Puppe Besitz des Mädchens ist (d.h. das Mädchen wird geschlagen, das die Puppe hat). Diese Ambiguität wird durch zwei parallel konstruierte MRS-Frames für das Wort *with* erfasst. In diesen verweist der angegebene Handle im einen Fall auf *h1* (der auf das Ereignis *hit* verweist), im anderen Fall auf *h5* (über diesen wird auf das Objekt *girl* verwiesen).

Map. Die interne Repräsentation der Satzbedeutung wird vor allem durch den Semantik-Formalismus bestimmt, der durch die Informationen der MRS Dependency Graphs sehr viel Struktur vorgibt. Das Netzwerk selbst leistet die Kompression dieser Informationen, die Abbildung der Abhängigkeiten zwischen den abgebildeten Frames (über deren Handles) sowie die selbstorganisierte Zuweisung der Frame Nodes als repräsentative Einheiten für MRS-Frames mit ähnlichen semantischen Eigenschaften.

Der Frame Selector (Abb. 4.3(c)) schließlich ist ein technisches Hilfsmittel, das dazu dient, die Aktivierungsstärke der Frame Nodes und damit zusammengehörige Aktivierungsmuster in der Frame Map zu kontrollieren. Er wird ebenfalls durch den Sequence Processor aktiviert, und bestimmt in der Folge für die Frame Map, welche Frame Nodes als Targets für den jeweils aktuell verarbeiteten Satz verwendet werden. Dabei ermöglicht er durch eine graduelle Auswahl der Frame Nodes die Modellierung psycholinguistischer Effekte wie etwa stereotypische Erwartungen semantischer Rollen und Eigenschaften, semantisches Priming oder Parallelverarbeitung bei Ambiguität. Vor allem aber erfüllt der Frame Selector den pragmatischen Zweck, durch die Auswahl der zum Input korrespondierenden Frames den Trainingsaufwand zu reduzieren. Das Training des Netzwerks ist sehr komplex organisiert und erfolgt neben unüberwachten Trainingsverfahren, etwa für das SARDNet sowie für die Frame Node Indicator Map, die die Entwicklung des Frame Selectors leitet, durch den Algorithmus Backpropagation Through Time. Der Fehler wird dabei von den Target-Vektoren am MRS Dependency Graph über die jeweils ausgewählten Frame Nodes und über den rekurrenten Hidden-Layer bis zum Input-Layer des Sequence Processors zurückpropagiert.<sup>9</sup>

Bei allgemeinen Performanztests nach strengen Evaluationskriterien, die die korrekte Aktivierung aller MRS-Frames für einen Satz erfordern, zeigt INSOMNet zwar verschiedentlich Schwächen bei der Genauigkeit der abzubildenden Argumentstruktur. In vielerlei Hinsicht zeigt das Modell aber eine beeindruckende Leistungsfähigkeit. So ist es innerhalb gewisser Grenzen tolerant gegenüber Problemen, die in natürlichsprachlichen Korpora durch Ungenauigkeiten entstehen, etwa durch ungrammatische, verkürzte oder dialektal gefärbte Äußerungen sowie Nebengeräusche. Zudem kann es erfolgreich verschiedene Formen von Ambiguität auflösen und ermöglicht diesbezüglich die Simulation verschiedener psycholinguistischer Effekte. Dabei weist INSOMNet ein kognitiv plausibles Verhalten wie Antizipation, Inkrementalität und Nichtmonotonität auf, das sich mit empirischen Beobachtungen deckt.

### 4.4. Direkte Assoziation und Integration

Eine ganze Reihe weiterer Modelle wählt einen wesentlich einfacheren Zugang zur Leistung und Kontrolle der Abbildung zwischen sprachlicher Form und Bedeutung. In diesen

---

<sup>9</sup>INSOMNet kann holistisch in einer Phase trainiert werden, allerdings stellt Mayberry (2003, S. 51) fest, dass bei einem in zwei Phasen gestuften Training (mit getrenntem Training für die selbstorganisierten und die mit Backpropagation trainierten Komponenten) eine etwas bessere Performanz erzielt wird.

wird in einer einfachen direkten assoziativen Zuweisung Form auf Bedeutung abgebildet<sup>10</sup>, wobei unterschiedliche Repräsentationsformate zur Konstruktion der Bedeutung verwendet werden. Diese rein assoziativen Modelle können am besten dadurch weiter unterteilt werden, dass zwischen Modellen unterschieden wird, die ausschließlich sprachliche Sequenzen auf statische Repräsentationen von Bedeutung abbilden (etwa S. L. Frank et al., 2003b, 2009; Dominey & Boucher, 2005), und integrativen Modellen, die Kombinationen verschiedener Input-Strukturen und auch verschiedener Transformationsaufgaben aufeinander und auf Target-Repräsentationen abbilden (etwa Mayberry et al., 2005, 2009; Dienes, Altmann & Gao, 1999; Altmann & Mirkovic, 2009). Zwar zeigen sich einerseits für diese rein assoziativ gesteuerten Modelle Beschränkungen für die Komplexität der sprachlichen Strukturen und der abbildbaren Bedeutungsstrukturen, andererseits zeigen diese Modelle andere interessante Phänomene und zielen auf andere Fragestellungen ab als die zuvor besprochenen auf das Skalierbarkeitsproblem fokussierten Modelle. Im Vordergrund der Besprechung in diesem Abschnitt stehen die verschiedenen Versionen des DSS-Modells (S. L. Frank et al., 2003b, 2009) und des CIA-Nets (Mayberry et al., 2005, 2009). Wir werden in Kapitel 9.3.2 nochmals auf einige Aspekte dieser Modelle zu sprechen kommen, wenn es darum gehen wird, situational bzw. sensomotorisch verankerte Satzverarbeitung zu modellieren.

##### **4.4.1. Direkte Assoziation: Distributed Situation Space**

Das Modell von S. L. Frank et al. (2003b), das wir im Folgenden vorstellen werden, unterscheidet sich von den vorausgehenden (und vielen weiteren Modellen, etwa Hadley et al., 2001; Mayberry et al., 2005; Dominey & Boucher, 2005; Desai, 2007) des Sprach- bzw. Textverstehens vor allem darin, dass die Verwendung von expliziten semantischen Enkodierungen mit Propositionen oder anderen symbolischen Formalisierungen vermieden wird. Sie stellen diesen Ansätzen das Distributed Situation Space-Modell (DSS-Modell) entgegen, in dem sich Ereignisse nur hinsichtlich ihrer Korrelation zueinander unterscheiden bzw. hinsichtlich möglicher Negationen, Konjunktionen und Disjunktionen. In den verschiedenen Versionen und Untersuchungen des Modells wurden hauptsächlich zwei Ziele verfolgt: Zum einen wird die Modellierung von wissensbasierten Inferenzen beim Verstehen von Texten forciert (S. L. Frank et al., 2003b, 2007, 2008) und dabei die Rolle von Weltwissen für die Diskursverarbeitung sowie die Form möglicher Implementierungen diskutiert. Zum anderen wird die semantische Systematizität in genuin konnektionistischen Modellen untersucht (S. L. Frank & Haselager, 2006; S. L. Frank et al., 2009). In diesen Modellierungen richten sich die Autoren direkt gegen die Behauptung von Fodor und Pylyshyn (1988), konnektionistische Modelle könnten nur dann Systematizität aufweisen, wenn sie ein klassisches Symbolsystem implementieren (vgl. Kap. 3.1.4). Diese Kritik betreffend benennen sie Aspekte, die für die konnektionistische Modellierung beachtet werden sollten, um sicherzustellen, dass die Systematizität der Performanz tatsächlich von inhärent konnektionistischen Eigenschaften und nicht von der Implementierung symbolischer Derivate herrührt.

---

<sup>10</sup>In dieser Hinsicht haben diese Modelle eine starke Ähnlichkeit zum frühen Ansatz von (McClelland & Kawamoto, 1986). In ihrem Modell werden statisch dargebotene Sätze auf eine propositionale Repräsentation abgebildet.

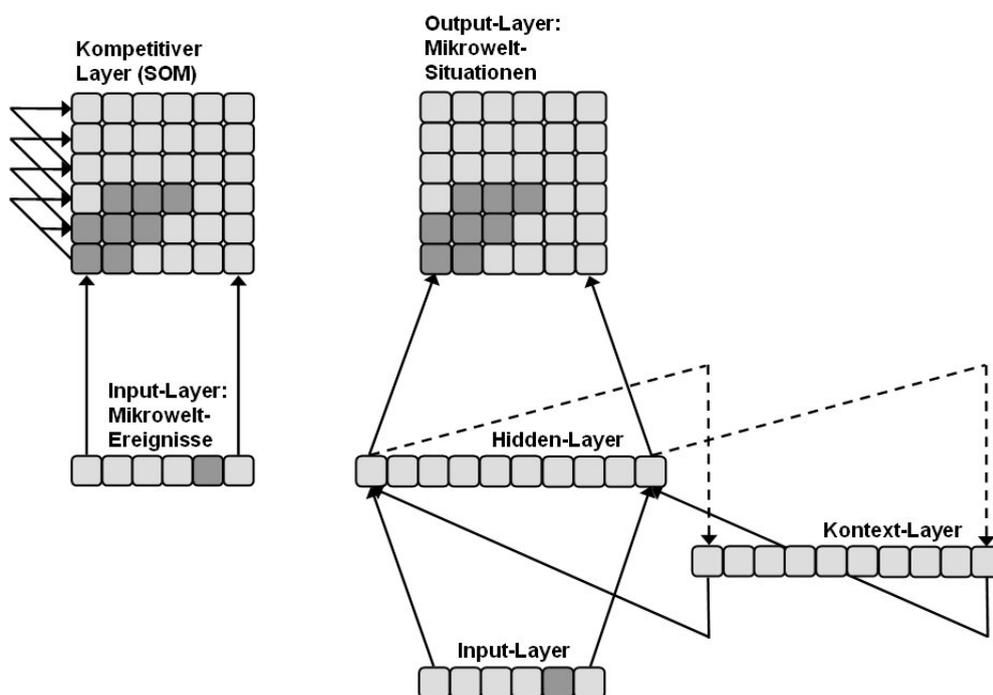


Abbildung 4.4.: Architektur des DSS-Modells (Frank et al., 2003): Anhand eines SOMs werden aus Basis-Events Situationsvektoren für einfache und komplexe Ereignisse einer Mikrowelt erzeugt. In einem zweiten Schritt werden mit einem SRN Sätze einer Mikrosprache auf diese Situationsvektoren abgebildet (vgl. z.B. Frank et al., 2007).

Das DSS-Modell wird in zwei Schritten trainiert (vgl. Abb. 4.4): In einem ersten Schritt werden mit einer SOM Klassifikationsstrukturen für semantische Basis-Repräsentationen ermittelt. Die dabei generierten Repräsentationen dienen als Situationsvektoren für die weitere Modellierung. In der späteren Version von S. L. Frank et al. (2009) werden über andere kompetitive Algorithmen einfachere Situationsformate (lineare Vektoren) generiert. Anhand boolescher Operationen werden komplexe Events als Konjunktionen und Disjunktionen von Basis-Ereignissen abgeleitet. Die Situationsvektoren dienen in einem zweiten Schritt als Targets für die sequenzielle Verarbeitung in einem SRN: Hier werden Sätze als Wortsequenzen auf die statischen Repräsentationen korrespondierender Situationen abgebildet. Diese Architektur und der Simulationsablauf sind vergleichsweise einfach strukturiert. Der eigentliche Kern der Leistung dieses Modells beruht auf den zugrundeliegenden Basis-Repräsentationen und ihren Kombinationen, auf der Generierung der Situationsvektoren für einfache und komplexe Ereignisse und Zustände sowie auf den Analysemethoden für die Evaluation der Satzverarbeitung der trainierten Netzwerke.

Das DSS-Modell basiert auf dem Ansatz von (Golden & Rumelhart, 1993) und erweitert deren Diskursverstehens-Modell. Die Grundidee von Frank und Kollegen ist, dass Ereignisse als Punkte und zeitliche Abfolgen von Ereignissen als Trajektorien in einem hochdimensionalen Ereignisraum repräsentiert sind. Die Situationsvektoren werden in einem kompetitiven Netzwerk generiert und bilden Beobachtungen von Ereignissen einer Mikrowelt ab. Solche Beobachtungen entsprechen Propositionen, die die möglichen Ereignisse der Mikrowelt beschreiben. Basis-Ereignisse sind etwa ES REGNET, PERSONEN

SPIELEN FUSSBALL, PERSON A GEWINNT, PERSON B IST MÜDE. Solche Ereignisse können auch mit einer bestimmten Wahrscheinlichkeit miteinander oder nacheinander erfolgen. Solche bedingten Wahrscheinlichkeiten für Ereignisse werden über Kombinationen von Basisereignissen anhand boolescher Operatoren für Negation, Konjunktion und Disjunktion dargestellt. Damit werden komplexe Ereignisse wie zum Beispiel PERSON A UND PERSON B SPIELE DRAUSSEN FUSSBALL, PERSON B GEWINNT ODER PERSON A GEWINNT abgebildet. Zur Erzeugung der nicht-propositionalen Situationsvektoren verwenden S. L. Frank et al. (2003b) lokalistisch enkodierte (und damit maximal unähnliche) Vektoren für Basis-Ereignisse<sup>11</sup> und bilden die Beobachtungen der Mikrowelt-Ereignisse in SOMs in topographischer Organisation ab. Interessant ist dabei folgende Eigenschaft: Je mehr Information über ein Ereignis gegeben ist, desto geringer fällt der benötigte Repräsentationsraum aus. Denn während Disjunktionen den Repräsentationsraum aller möglichen Ereignisse benötigen, genügt für die Abbildung von Konjunktionen die Schnittmenge des Repräsentationsraums der beteiligten Ereignisse. Die generierten Situationsvektoren verfügen über implizite Informationen zu den Constraints, die in der beschriebenen Mikrowelt herrschen. So spielen etwa die beschriebenen Personen niemals drinnen Fußball, und sie spielen nur dann draußen, wenn es nicht regnet. Bedeutung entspricht in dieser Konzeption der Konstruktion einer internen Situationsrepräsentation im Sinne eines mentalen Situationsmodells, das die aktuellen Zustände der Welt und deren Zusammenhänge erfasst. Anders als in den zuvor beschriebenen Modellen werden hier nicht die Repräsentationen eines internen Layers als Kern der Bedeutung verstanden, sondern der Situationsvektor am Output-Layer, der die Verortung des Ereignisses im verteilten Situationsraum wiedergibt.

Anhand eines SRNs werden inkrementell dargebotene Sätze einer Mikrosprache, mit der die Ereignisse des beschränkten Ereignisraums der Mikrowelt beschrieben werden können, auf korrespondierende Situationsvektoren abgebildet. Zur Verarbeitung der Sätze muss das Modell die korrekten Situationsvektoren abbilden. Dabei zeigten die Simulationen für die frühe Version des Modells, dass dieses sowohl neue Sätze für bekannte Ereignisse (Beschreibung eines im Trainingsset enthaltenen Situationsvektors mit einer für diesen nicht verwendeten Satzstruktur) als auch Sätze für gänzlich neue Ereignisse (komplexe Ereignisse, die nicht im Trainingsset enthalten waren) korrekt interpretieren konnte. Zur erfolgreichen Verarbeitung musste das DSS-Modell zudem auf Basis der impliziten Informationen über situationale Constraints Inferenzen leisten.<sup>12</sup> Aus der Performanz können auch verschiedene Gedächtniseffekte abgeleitet werden, etwa der kognitive Aufwand zur Integration eines neuen Ereignisses in eine Geschichte, oder die relative Wahrscheinlichkeit, dass ein Ereignis im Verlauf der Verarbeitung wieder vergessen wird. In der neuen Version des Modells (S. L. Frank et al., 2009) wurde das Universum möglicher Ereignisse der Mikrowelt deutlich erweitert und die Tests zur Systematizität des Modells bei der Verarbeitung von Sätzen bzw. der Generierung korrespondierender Situationsvektoren sys-

<sup>11</sup>Auf diese Weise bieten die Input-Vektoren keinerlei Informationen über den Inhalt der Events oder über seine Zusammenhänge und Ähnlichkeiten zu anderen Events, denn durch die lokalistische Enkodierung sind sie vollkommen arbiträr und orthogonal zueinander.

<sup>12</sup>Beispielsweise können Präsuppositionen zu zwingenden vorausgehenden Ereignissen oder Zuständen geschlossen werden: Wenn eine der zwei Personen der Mikrowelt in einem Spiel gewinnt, müssen zum Zeitpunkt zuvor beide Personen am selben Ort (in der Mikrowelt entweder drinnen oder draußen) gewesen sein. Beispielsweise können auch wahrscheinliche Ergebnisse zukünftiger Ereignisse geschlossen werden: Wenn die beiden Personen ein Spiel spielen und eine der Personen zum vorangegangenen Zeitpunkt müde war, so ist es wahrscheinlicher, dass diese Person das Spiel verlieren wird.

tematisiert. Dabei wurde eine Hierarchie verschiedener Systematizitäts-Stufen entwickelt, die über die Generalisierungsfähigkeit auf syntaktischer Ebene auch die Leistungen auf semantischer Ebene bewertet. Die im Verhältnis zur jeweiligen Komplexitätsstufe gute Performanz des Modells stützt die Kritik der Autoren an der Annahme von Fodor und Pylyshyn (1988). Allerdings bestätigen sie auch, dass die Systematizität der Verarbeitung mitunter auf unsauberen Aktivierungen beruht (etwa hinsichtlich der Koaktivierung falscher oder unbeteiligter Events), sowie dass das von G. F. Marcus (1998) geäußerte Problem der Generalisierung außerhalb des Trainingsereignisraums (vgl. Kap. 3.1.4.1) mit ihrem Modell weiterhin ungelöst bleibt.

**Weitere Modelle** Nur kurz erwähnt werden sollen zwei weitere Modelle, die ebenfalls eine direkte Abbildung von sequenziell dargebotenen Sätzen auf einen statischen Input leisten: Zum einen ein hybrides Modell von Dominey und Boucher (2005), in dem Sprachverstehen durch ein konnektionistisches Modul geleistet wird, zum anderen ein konnektionistisches Modell von (Desai, 2007). Beide Modelle beschäftigen sich speziell mit dem Sprach- und Bedeutungserwerb. Dementsprechend ist die abzubildende Komplexität eher gering. Im Modell von Dominey und Boucher erwirbt das System sprachliches Wissen anhand sprachlich beschriebener Videoereignisse, in denen Ereignisse physischen Kontakts und einfache räumliche Relationen zwischen einfachen Objekten dargestellt werden (vgl. auch Dominey, 2003). Die Vorverarbeitung generiert eine propositionale Repräsentation der Szenen und eine statische Repräsentation des beschreibenden Satzes. Durch die konnektionistische Architektur werden über grammatische Schemata Prinzipien der Konstruktionsgrammatik implementiert. Dabei erfolgen zwei Abbildungen: einerseits von Wörtern auf individuelle Komponenten eines Ereignisses, andererseits von grammatischen Funktionen auf funktionale Rollen der Komponenten in den Ereignissen. Die erworbenen Repräsentationen ermöglichen eine Generalisierung zum Verstehen bislang nicht gehörter Sätze. Ein Großteil der Leistung wird allerdings durch die Vorverarbeitung der anderen Module geleistet.

Im Modell von Desai (2007) werden syntaktische und semantische Bootstrapping-Effekte und ihr Einfluss auf den Vokabelerwerb untersucht (zu einer frühen Version des Modells vgl. Desai, 2001). Die Architektur besteht aus einem erweiterten SRN, dessen Output-Layer durch einen zweiten Kontext-Layer ebenfalls rekurrent gestaltet wurde und zudem eine direkte Verbindung aus einem vorgeschalteten nicht-rekurrenten Hidden-Layer erhält. Im Modell werden Sätze Wort für Wort auf eine korrespondierende propositionale Repräsentation abgebildet. Diese stellt durch Slots für zwei Objekte, eine Handlung und eine Kausalitätsbeziehung eine sehr beschränkte visuelle Szene dar. Das Modell bietet einen Erklärungsansatz für Compliance-Effekte<sup>13</sup> und für die vom Erwerbsalter abhängigen Veränderungen der Beeinflussung – weg vom syntaktischen Rahmen und hin zur semantischen Information als Interpretationsbasis. Die Effekte werden durch konkurrierende Hinweisreize für die Vorhersage der Kausalität erklärt.

---

<sup>13</sup>Compliance-Effekte entstehen, wenn die Interpretation eines Satzes entweder eher durch den syntaktischen Rahmen oder eher durch die Bedeutung des Verbs geleitet wird. Dies führt in der frühen Erwerbsphase zum Beispiel dazu, dass Verben in einem falschen syntaktischen Rahmen eher akzeptiert werden (vgl. Desai, 2007, S. 2273f).

#### 4.4.2. Assoziation und Integration: Coordinated Interplay Account

Die folgenden Modelle unterscheiden sich insofern von den eben vorgestellten, als dass sie Informationen aus verschiedenen Input-Quellen integriert auf eine oder mehrere Ziel-Strukturen abbilden. Anders als im DSS-Modell muss also nicht nur sprachlicher Input verarbeitet werden, sondern beispielsweise auch visueller Input. Dabei geht es vor allem darum, für die zu leistenden Aufgaben die gegenseitige Ergänzung und Interaktion verschiedener Informationsquellen zu nutzen.

Im Falle des Coordinated Interplay Account Network (CIA-Net, Mayberry et al., 2009, 2005) geht es darum, die Nutzung des visuellen Kontexts zur Interpretation und Disambiguierung von Sätzen zu simulieren. Damit werden im Unterschied zu vielen anderen Modellen nicht Lesezeiten oder Verstehensleistungen modelliert, sondern visuelle Aufmerksamkeit und Antizipation – speziell zu Daten aus Visual-World-Experimenten, in denen Blickbewegungen auf visuellen Szenen zu Sprachverarbeitungsprozessen in Beziehung gesetzt werden (Cooper, 1974; Tanenhaus, Spivey-Knowlton, Eberhard & Sedivy, 1995, siehe auch Kap. 5.2.2). Ausgangspunkt bzw. Motivation von CIA-Net ist sein theoretisches und empirisches Gegenstück: der Coordinated Interplay Account (u.a. Knoeferle & Crocker, 2006), der die Wechselwirkung von visueller Szene, sprachlicher Äußerung und Weltwissen bei der Sprachverarbeitung beschreibt. Diesem Ansatz zufolge wird die zeitliche Abfolge der Satzverarbeitung bei korrespondierendem visuellem Kontext durch zwei sukzessive Prozesse geleitet: (a) zuerst durch die sprachlich gestützte Aufmerksamkeitslenkung zur Identifikation beteiligter Objekte und Ereignisse und die Antizipation möglicher Referenten, (b) anschließend durch die Beeinflussung der Interpretation der Äußerung über die Informationen aus dem visuellen Kontext. Solche Antizipations- und Inferenz-Prozesse sollen mit CIA-Net simuliert werden.

Die Architektur von CIA-Net (vgl. Abb. 4.5, in der neueren Variante von Mayberry et al. (2009)) besteht aus einem erweiterten SRN, das zunächst einmal zwei verschiedene Inputformate im rekurrenten Hidden-Layer integriert. Am sprachlichen Input-Layer werden Wort für Wort Sätze dargeboten. Am visuellen Input-Layer wird eine komplexe visuelle Szene dargeboten, die drei Personen in zwei unterschiedlichen Handlungen abbildet: Die Personen 1 und 2 sind in Handlung A involviert, die Personen 2 und 3 in Handlung B (z.B. DER PIRAT WÄSCHT DIE PRINZESSIN und DIE PRINZESSIN MALT DEN FECHTER). Die visuelle Szene wird damit in einer Slot-Filler-Repräsentation für zwei gleichzeitig zutreffende Propositionen dargestellt. Das gewählte Format beschränkt natürlich stark die Variabilität der darstellbaren Szenen, ist aber direkt dem Aufbau der empirischen Studien geschuldet, die mit CIA-Net modelliert werden sollen. Über den rekurrenten Hidden-Layer werden die Informationen integriert und auf den Output-Layer abgebildet, der eine statische propositionale Repräsentation mit Zuweisung thematischer Rollen von einer der beiden möglichen Handlungen vorgibt. Die zweite Komponente, um die das SRN erweitert wurde, simuliert visuelle Aufmerksamkeit (covert attention). Aufmerksamkeit ist als Gating-Vektor implementiert, der die Aktivierungsstärke zwischen verschiedenen Slots des visuellen Input-Layers verschiebt. Diese Aufmerksamkeitsverschiebung wird während der Trainingsphase explizit gelernt und in der Testphase dazu verwendet, die Belegung der thematischen Rollen zu antizipieren.

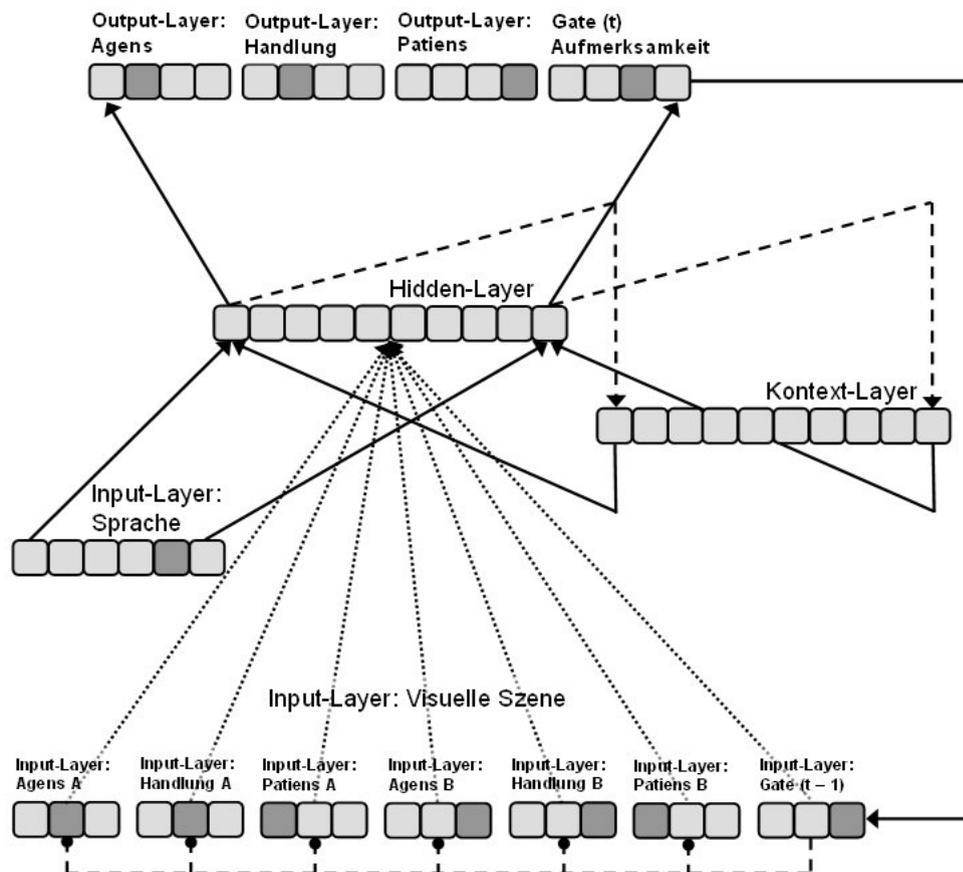


Abbildung 4.5.: Die Architektur von CIA-Net besteht aus einem erweiterten SRN (vgl. Mayberry et al., 2009, S. 460). Neben einem sequenziellen sprachlichen Input erhält der Hidden-Layer Input aus einer Slot-Filler-Repräsentation einer visuellen Szene, die zwei Propositionen enthält. Im Hidden-Layer werden die Input-Informationen komprimiert und integriert, um eine Abbildung auf eine Slot-Filler-Repräsentation einer einfachen Proposition zu leisten.

Das Modell erfasst erfolgreich Befunde aus mehreren empirischen Studien im Visual-World-Paradigma, etwa die Antizipation von Rollenbelegungen durch frühe linguistische Hinweise und folgende Inferenzen anhand des visuellen Kontexts, oder die Disambiguierung lokal ambiger Sätze durch die visuellen Stimuli. Dabei zeigt es die von den Autoren geforderten zentralen kognitiven Charakteristiken Inkrementalität, Antizipation, Adaptivität sowie Integration und zeitliche Koordination verschiedener Informationsquellen. Neben der Modellierung dieser spezifischen Aspekte der Performanz kann CIA-Net als allgemeines Modell der situierten Sprachverarbeitung und des situational verankerten Spracherwerbs verstanden werden. Allerdings äußern sich die Autoren nicht dazu, welche Komponente des Modells als Substrat der Repräsentation von Bedeutung zu verstehen ist.

**Weitere Modelle** Abschließend wollen wir noch das Modell von Dienes et al. (1999) und das darauf beruhende theoretische Modell von Altmann und Kamide (2009) erwähnen. Die grundlegenden Annahmen zur Sprachverarbeitung und zur Integration von Informationen verschiedener Quellen bzw. Domänen decken sich stark mit den Ansätzen von Mayberry et al. (2009). Im Modell von Dienes et al. (1999) wird die domänenübergreifende Abbildung von Informationen und der Transfers von implizitem Wissen zwischen verschiedenen Domänen simuliert. Es geht also zunächst nicht um die Modellierung von Sprachverstehen. Im Zentrum steht die Übertragung von Wissensstrukturen einer Domäne auf eine andere Domäne, in der dieses Wissen nicht explizit erlernt wurde. Das Modell basiert auf der SRN-Architektur, mit dem einzigen relevanten Unterschied, dass es über zwei Input- und zwei Output-Layer verfügt. Diese implementieren jeweils die Verarbeitungsrouten zweier unterschiedlicher Domänen und integrieren diese in einem vorgeschalteten Hidden-Layer und im rekurrenten Hidden-Layer. Als Beispiel wird das Experimentalparadigma Artificial Grammar Learning herangezogen: In diesem kann gezeigt werden, wie übertragbares sequenzielles Wissen beispielsweise von der Domäne Buchstaben (Graphemsequenzen) auf die Domäne Töne (Tonfolgen) übertragen und dort ohne weitere Übung angewendet werden kann. In Simulationen konnten eine Reihe von Effekten repliziert werden, die empirisch für diese Aufgabe gezeigt wurden. Die Fähigkeit von SRNs, Informationen aus mehreren Domänen zu integrieren und darüber hinausgehend aus der Interaktion dieser Domänen versteckte statistische Regularitäten abzuleiten, die in den isolierten Domänen nicht zugänglich sind, wurde auch in Simulationen von Christiansen et al. (1998) und Christiansen et al. (2005) gezeigt. Altmann und Mirkovic (2009) erweitern die Annahmen des Modells von Dienes et al. (1999) zu einem allgemeinen Ansatz der menschlichen Sprachverarbeitung und der Integration von Informationen aus verschiedenen Domänen – etwa der sprachlichen und der visuellen Verarbeitung. Sie leiten daraus zentrale Prinzipien der Antizipation bei der Sprachverarbeitung ab. Satzverstehen wird dabei als Abbildung zwischen sich entwickelnder Satzstruktur und situationaler Ereignisrepräsentation verstanden, wobei Vorhersage-Prozesse gleichermaßen durch den linguistischen und den nicht-linguistischen (z.B. visuellen) Input sowie durch gespeichertes (top-down) Wissen geleitet werden. In diesem Zusammenhang diskutieren die Autoren in einem umfassenden Überblick empirische und komputationale Studien, die die Interaktion solcher antizipatorischen und integrativen Prozesse illustrieren.

## 4.5. Konzeption semantischer Verarbeitung

Fassen wir zusammen: Welche Konzeption von Bedeutung und Verstehen wird in konnektionistischen Modellen der PDP-Tradition etabliert und wie wird diese in Modellen der Satzverarbeitung umgesetzt? Und: Gibt es überhaupt Überschneidungen zwischen den verschiedenen Ansätzen, anhand derer man eine gemeinsame Konzeption identifizieren kann? Wählen wir als Ausgangspunkt die grundlegenden Derivate der Bedeutung im Netzwerk: Wo und wie wird Bedeutung in konnektionistischen Netzwerken gespeichert? Hierzu haben wir in Kapitel 2.2.2 verschiedene Positionen kennengelernt. Es wurde zwischen lokalistischen und verteilten Repräsentationen unterschieden. Bei lokalistischen Repräsentationen entspricht die Aktivierung einzelner bedeutungstragender Units dem Abruf von konzeptuellem Wissen. Bei verteilten Repräsentationen entspricht die Aktivierung mehrerer Units (d.h. eines spezifischen Aktivierungsmusters) dem Abruf von konzeptuellem Wissen. Während dabei zumeist Aktivierungsmuster innerhalb eines Hidden-Layers als Repräsentationsort verstanden werden, sind in eher neurofunktional ausgerichteten Ansätzen (z.B. Pulvermüller, 1999; Wennekers, Garagnani & Pulvermüller, 2006; Garagnani, Wennekers & Pulvermüller, 2008) Aktivierungsmuster gemeint, die im Sinne hebbischer Cell-Assemblies über das gesamte Netzwerk verteilt sind. Das eigentliche, langfristig gespeicherte Wissen kann den Gewichtsstrukturen zugewiesen werden, über die die spezifischen Aktivierungsmuster generiert werden. Eine mögliche Auslegung zum Verhältnis zwischen Aktivierungsmustern und Gewichtsstrukturen ist, dass das Wissen über Konzepte in den Aktivierungsmustern gegeben ist, das Wissen über konzeptuelle Relationen aber in den Gewichten gespeichert wird (vgl. Hinton, 1986; Rogers & McClelland, 2003; McClelland & Rogers, 2003). Eine andere Möglichkeit ist, die Gewichtsstrukturen als Repräsentationsort des indifferenten Gesamtwissens zu verstehen und einzelne unterscheidbare Wissensinhalte oder Interpretationen in den dadurch hervorgerufenen Aktivierungsmustern zu sehen. Interpretiert man die Aktivierungsmuster als Vektoren, so können die Bedeutungen als Verortungen in einem hochdimensionalen Zustandsraum verstanden werden, in dem die semantischen Strukturen und Zusammenhänge der repräsentierten Außenwelt über räumliche Approximation und Distanz der Bedeutungsverortungen erfasst werden (vgl. z.B. Laakso & Cottrell, 2000; Bullinaria, 1997; Elman, 2004). Was ist in konnektionistischen Netzwerken Verstehen und Interpretation? Als Interpretation kann jede geleistete Transformation von einem Input-Muster zu einem Output-Muster verstanden werden. Dabei entspricht der Verstehensprozess der Transformation, die als Constraint-Satisfaction-Prozess durch die internen Strukturen (interne Repräsentationen in Gewichten und Aktivierungsmustern) erfolgt.

Überblickt man die verschiedenen Konzeptionen von Bedeutung, Wissen und Verstehen, die in der konnektionistischen Literatur geäußert werden, so ergibt sich ein eher uneinheitliches Bild, das viel Raum für Missverständnisse und Unschärfen bietet. Gerade was die Konzeption von semantischem Wissen angeht, haben wir in den letzten beiden Kapiteln 3 und 4 sehr unterschiedliche Ansätze kennengelernt, in denen mit derselben zugrundegelegten Architektur der SRNs Bedeutung erfasst wurde: Zum einen erfolgte dies über die Kollokationen von Wörtern in Sätzen, über die sich im Hidden-Layer spezifische semantische Gliederungen im Zustandsraum ergeben. Das konzeptuelle Wissen ist hier also in den internen Repräsentationen zu suchen. Dass eine solche rein intensionale Konzeption

von Sprachverstehen aber keine endgültig befriedigende Lösung sein kann, wird zum Beispiel bereits von (Elman, 1991) selbst konstatiert. Zum anderen erfolgte die Konzeption von Bedeutung über die Referenz, d.h. über die referenzielle Abbildung struktureller Zusammenhänge in der Außenwelt. Allerdings wird in den meisten Fällen nicht explizit geäußert, welche Netzwerk-Komponente als Derivat der Bedeutung verstanden wird. Dies birgt eine weitere Quelle für Unschärfen: Sind die Aktivierungsmuster der Output-Layer bereits Bestandteil des Wissens des Netzwerks? Sind diese Muster also bereits als kognitive Repräsentationsstrukturen zu verstehen? Dies würde in etwa der Konzeption von Rogers und McClelland (2003) entsprechen, in der die Aktivierungsmuster (der Output-Layer) das Wissen über Konzepte repräsentieren, während das Wissen über konzeptuelle Relationen in den internen Strukturen repräsentiert wird. Oder sind sie nur Strukturen der Außenwelt, etwa sensorische Reize, die ein Teil der Erfahrung des Netzwerks werden und die kognitiven Repräsentationsstrukturen prägen, die sich in den internen Strukturen entwickeln? Wird also netzwerkintern eine Abbildung der gesamten Situation einschließlich Konzepten und Konzept-Relationen repräsentiert? Oder aber: Sind mentale Konzepte ein Konstrukt, deren Annahme mithilfe des Konnektionismus verworfen werden kann, da in diesem nur noch konzeptuelle Relationen gegeben sein müssen, die die richtige Reaktion auf einen Umweltreiz erlauben (vgl. Peschl, 1996)?

Im SG-Modell (McClelland et al., 1989; St. John, 1992) sowie im CSCP-Modell (Rohde, 2002) werden über zentrale Hidden-Layer selbstorganisierte interne Strukturen entwickelt, die die Geschehnisse in der Welt abbilden (vgl. Kap. 4.2). Das gesamte konzeptuelle Wissen einschließlich seiner Relationen ist also in internen Repräsentationen im Sinne eines mentalen Modells der Welt erfasst. Satzverstehen entspricht hier der Konstruktion eines mentalen Modells des relevanten Weltausschnitts. Während dies unseres Erachtens eine erstrebenswerte Konzeption des Satzverstehens und der semantischen Bedeutung von Sätzen ist, bergen die Ansätze zwei Probleme in sich. Erstens wird die selbstorganisierte Entwicklung der Repräsentationsstrukturen durch einen Mechanismus erreicht, dessen psychologische Realität fragwürdig ist: Die kognitive Enkodierung von Wissen anhand eines Abfrage-Mechanismus für Teilstrukturen der jeweils erlernten Wissensinhalte erscheint nicht plausibel und deckt sich mit keinem uns bekannten Befund der Lernpsychologie. Wird der Mechanismus im Modell aber nur als Mittel zum Zweck der Modellierung verstanden, der eine selbständige Enkodierung und Komprimierung des Inputs ermöglicht, so eröffnet das Modell keinen Erklärungsansatz für den Erwerb von Bedeutung und die Kontrolle der Bedeutungskonstruktion. Zweitens liegen den intern entwickelten Wissensstrukturen propositionale Repräsentationen zugrunde. Propositionen als elementare Bausteine mentaler Strukturen der Bedeutung von Texten und Diskursen haben eine lange Tradition in der Kognitionspsychologie und die Annahme ihrer psychologischen Realität wurden durch Befunde aus empirischen Studien unterstützt (für eine umfassende Darstellung der diesbezüglichen Ergebnisse vgl. Kintsch, 1998). Allerdings sind Propositionen keine Strukturen, die wir direkt in der Welt beobachten können. Vielmehr sind sie eine formalisierte Darstellung der Ergebnisse kognitiver Verarbeitung, in denen die Zusammenhänge der Dinge und Ereignisse in der Welt bereits enthalten sind. Dem kognitiven System diese Strukturen als Grundlage der Wahrnehmung der Welt zu geben, ist unrealistisch und nimmt dem Modell durch die explizite Darbietung der relevanten Strukturen und Zusammenhänge die Aufgaben ab, die es eigentlich selbst erfüllen soll. Taatgen und Anderson (2008, S. 4) bemerken hierzu in einem allgemeinen Statement zur kognitiven Modellierung: „A model

should learn its own task-specific knowledge. Building knowledge into a model increases its specificity, and may decrease its validity.“ Genau dieses Einbauen von explizitem Wissen geschieht unseres Erachtens aber mit der Verwendung von Propositionen. Dies wird ausführlich von Waskan (2001) in einer Diskussion der Prägung und Verzerrung der Performanz konnektionistischer Modelle durch die Darbietung inadäquater semantischer Input- und Ziel-Strukturen problematisiert.<sup>14</sup> Die Verwendung propositionaler Strukturen legt außerdem den Verdacht der Implementierung eines klassisch-symbolischen Systems nahe (vgl. Aizawa, 1997; S. L. Frank et al., 2009), da das Modell damit abstrakte diskrete Wissenseinheiten verwendet.

Im INSOMNet (Mayberry, 2003) erfolgt die Enkodierung zumindest für die Zuordnung der relevanten Frame Nodes zu den jeweiligen MRS-Frames ebenfalls selbstorganisiert durch SOMs, wodurch sich ähnlichkeitsbasierte Zuordnungen für die MRS-Frames ergeben. Die Repräsentation innerhalb der Frame Nodes erfolgt aber in der direkten Abbildung des jeweils anliegenden Wortes über den gewählten Frame Node auf den aktuellen MRS-Frame. Damit ist zwar das gesamte semantische Wissen der Situation (sowohl Konzepte als auch deren Relationen) intern repräsentiert. Dieses wird aber unmittelbar durch die zugrundegelegte semantische Repräsentation der MRS bedingt, die unseres Erachtens noch mächtiger und strukturell expliziter ist als die für das CSCP-Modell gewählte propositionale Repräsentation. Insofern gilt auch hier die zuvor geäußerte Problematisierung. Noch stärker trifft dies auf das CIA-Net (Mayberry et al., 2005, 2009) zu, in dem keine unabhängige interne Repräsentation der propositionalen Struktur forciert wird. Der Hidden-Layer bildet das semantische Wissen der Relationen innerhalb und zwischen verschiedenen Domänen in Beziehung zu den Output-Strukturen ab. Durch die direkte Abbildung auf die propositionale Struktur sind diese internen Strukturen stark durch die explizit vorgegebenen Informationen geprägt. Zudem ermöglicht diese Umsetzung keine Komprimierung bzw. Speicherung mehrerer gleichzeitig gegebener propositionaler Zustände. Das Modell kann also nur eine Proposition pro Äußerung verstehen. Ebenfalls problematisch ist die Implementierung eines expliziten Aufmerksamkeits-Mechanismus, der stark an eine Kontrollstruktur klassisch-symbolischer Modelle erinnert.

Wiederum selbstorganisiert erfolgt die Enkodierung im DSS-Modell (S. L. Frank et al., 2003b, 2009), in der die Ähnlichkeit verschiedener Ereignisse hinsichtlich der Konjunktion und Disjunktion ihres Auftretens erfasst wird. Im Unterschied zu den zuvor genannten Modellen verwenden Frank und Kollegen explizit nicht-propositionale Repräsentationen, um die damit verbundenen Probleme zu umgehen. Zwar basieren auch die erzeugten Situationsvektoren auf propositional interpretierten Basis-Ereignissen. Bei diesen werden jedoch keine expliziten strukturellen Informationen vorgegeben, da die Basisereignisse lokalistisch enkodiert sind und damit ein Minimum an strukturellen Informationen enthalten. Umgekehrt muss man sich bei diesem Repräsentationsformat allerdings fragen, inwiefern hier überhaupt Strukturen enthalten sind, aus denen konzeptuelles Wissen induziert werden kann. Vielmehr scheint es so, als ob die Situationsvektoren ausschließlich Wissen über konzeptuelle Relationen enthalten. Allerdings zeigen Frank und Kollegen auch Ansätze zur Auflösung pronominaler Referenz in diesem Modell, was ein indirektes

---

<sup>14</sup>Wir werden in Kapitel 9.3.2.2 nochmals auf den beträchtlichen Anteil zu sprechen kommen, den eine solche Vorgabe expliziter semantischer Strukturen an der Performanz und den Verarbeitungsmechanismen des Modells haben dürfte.

Vorhandensein konzeptuellen Wissens nahelegt.<sup>15</sup> Ein weiteres Problem des DSS-Modells ist die Trennung des Erwerbs der Situationsvektoren und des Erwerbs der darauf bezogenen sprachlichen Strukturen. Die Situationsvektoren werden explizit als mentale Strukturen im Sinne mentaler Situationsmodelle verstanden (vgl. S. L. Frank et al., 2009, S. 359) und damit die Aktivierungsmuster des Output-Layers als mentale Repräsentation aufgefasst. In dieser Form der Modellierung würde das aber bedeuten, dass sprachliches Wissen auf komplett unabhängig davon erworbenes und semantisch vollständiges Wissen abgebildet wird, und nicht auf die Struktur des konzeptuellen Wissens einwirken kann. Dies ist auch im CSCP-Modell der Fall, bei dem die Verstehens-Komponente und die Sprach-Komponente sukzessive trainiert werden. Wir bezweifeln, dass dies eine entwicklungspsychologisch realistische Perspektive ist.

---

<sup>15</sup>Wir vermuten, dass dies durch eine Verschmelzung von Relationen und Konzepten zustande kommt: Der Situationsvektor PERSON A GEWINNT enthält zugleich die Konzepte PERSON A und GEWINNEN sowie die Relation dieser beiden, nach der besagte Person als Sieger aus einer Begegnung hervorgeht. Ob eine solche Verschmelzung allerdings eine psychologisch plausible Konzeptualisierung semantischer Repräsentationen ist, ist zu hinterfragen.



## **Teil II.**

# **Koreferenz und lokale Kohärenz**



## 5. Anaphernaauflösung und lokale syntaktische Kohärenzen

Der zweite Teil dieser Arbeit beschäftigt sich mit der Erforschung der Satzverarbeitung und der Auflösung anaphorischer Referenzen anhand empirischer Studien und Modellierung. In diesem ersten Kapitel beleuchten wir einführend zu den beiden nachfolgenden Kapiteln die psycholinguistischen Hintergründe, die unserem Arbeitsfeld zugrunde liegen. Dabei werden wir in aller Kürze Theorien der syntaktischen Verarbeitung skizzieren, so dass wir die folgenden Darstellungen in einen größeren Kontext einbetten können. Anschließend werden wir eingehender die in Kapitel 3.2.3 erwähnte psycholinguistische Erforschung des Phänomens lokaler syntaktischer Kohärenzen (LSK) behandeln. Der Befund, dass lokale Subsequenzen unter temporärer Verletzung des globalen Zusammenhangs verarbeitet werden, hat Konsequenzen für gegenwärtige theoretische Ansätze der Sprachverarbeitung: Die zentrale Annahme, dass die Verarbeitung von Sätzen global Konsistent erfolgt, kann nur schwer damit in Übereinstimmung gebracht werden. Des Weiteren werden wir auf die Verarbeitung von Anaphern eingehen. Nach ersten Details zu strukturtheoretischen Annahmen in Form rein syntaktischer Constraints zur Auflösung anaphorischer Referenz besprechen wir empirische Befunde zur Online-Verarbeitung von Anaphern. Dabei werden auch Einschränkungen rein struktureller Ansätze gezeigt und Eigenschaften des zeitlichen Verlaufs sowie Einflüsse des syntaktischen und semantischen Kontexts erläutert.

### 5.1. Satzverarbeitung: Perspektiven und Modelle

Die Psycholinguistik beschäftigt sich mit der menschlichen Fähigkeit, Sprache zu verstehen, zu produzieren und zu erwerben. Dabei soll zum einen die mentale Repräsentation sprachlichen Wissens (u.a. lexikalisch-semantisches sowie grammatisches Wissen) identifiziert werden. Zum anderen sollen die kognitiven Prozesse erklärt werden, die der tatsächlichen Verarbeitung von Sprache zugrunde liegen und die die spezifischen sprachlichen Verhaltensmerkmale verantworten.<sup>1</sup>

Natürliche Sprache als unabhängig von seiner Verarbeitung beobachtbarer Gegenstand und die beobachtbaren Merkmale unseres sprachlichen Verhaltens weisen eine Reihe genereller Merkmale auf. Hierzu gehören unter anderem die folgenden Feststellungen: Sprache ist ein konventionalisiertes System arbiträrer Zeichen (vgl. de Saussure, 1967), das die

---

<sup>1</sup>Traditionell wird in diesem Zusammenhang zwischen einer kompetenzbasierten und einer performanzbasierten Perspektive unterschieden (vgl. Hemforth & Konieczny, 2002). Allerdings wird gerade diese Distinktion in neueren sprachtheoretischen Ansätzen angezweifelt, so auch im Konnektionismus, aus dessen Perspektive die Grenzen zwischen Repräsentation und Verarbeitung im klassischen Sinne nicht aufrecht erhalten werden können.

kontextungebundene Übertragung von Bedeutung ermöglicht. Sprache ist strukturiert – die Anordnung ihrer Grundeinheiten folgt einem System von Regularitäten (ein elementarer Unterschied zur Tiersprache, vgl. Tomasello, 2006). Sprache ist produktiv (vgl. Chomsky, 1957). Die durch einen Satz übertragene Bedeutung ist über deren kompositionelle Bedeutung (vgl. Kompositionalitätsprinzip nach Frege, 1966) hinaus abhängig von Constraints auf verschiedenen Ebenen, etwa Prosodie, Pragmatik oder situationalem Kontext. Die Verarbeitung von Sprache erfolgt unmittelbar und inkrementell (vgl. u.a. Altmann, 1999; Tanenhaus et al., 1995; Hemforth, Konieczny & Strube, 1993; Konieczny, 1996; Sedivy, Tanenhaus, Chambers & Carlson, 1999). Natürliche Sprache ist auf verschiedenen Ebenen ambig, die Mehrdeutigkeit bereitet uns in der Regel aber keine Schwierigkeiten. Sprachverarbeitung ist zudem robust gegenüber unvollständigen und verrauchten Informationen sowie gegenüber syntaktischen und semantischen Anomalien (Verletzungen der Wohlgeformtheit). Sprachverarbeitung ist sensitiv gegenüber Frequenzen und distributionalen Eigenschaften der Sprache (vgl. z.B. MacDonald, 1999).

**Zentrale Fragen** Warum weist die Sprache und ihre Verarbeitung diese Merkmale auf und welche kognitiven Mechanismen befähigen uns zur Verarbeitung? Details der Verarbeitung von Sprache und die genaue Ausprägung der oben genannten Merkmale sind Gegenstand grundlegender Debatten und Positionen. Für die Verarbeitung von Sätzen stehen dabei folgende Fragen im Zentrum (vgl. z.B. Hemforth & Konieczny, 2002; Crocker, Pickering & Clifton, 2000; Mitchell, 1994):

- Die Frage des zugrundeliegenden Mechanismus: Welche Mechanismen werden zur Konstruktion syntaktischer Strukturen bzw. zur Analyse syntaktischer Abhängigkeiten genutzt?
- Die Frage der Nutzung und Interaktion von Informationen: Zu welchem Zeitpunkt und in welcher Reihenfolge kommt welche Information (u.a. syntaktische, prosodische, lexikalisch-semantische, kontextuelle Information) zum Tragen? Inwiefern interagieren unterschiedliche Informationssysteme miteinander? Und auf welche Weise erfolgt zu welchem Zeitpunkt der Abgleich von datenbasierten (bottom-up) und wissensbasierten (top-down) Informationen?
- Die Frage der Repräsentation bei unsicherem Wissen: Erfolgt die Verarbeitung mehrdeutiger Informationen parallel (Aufrechterhaltung mehrerer oder sogar aller Interpretationsalternativen), seriell (Auswahl von nur einer Interpretationsalternative) oder über unvollständige Analyse (Aufschub der Verarbeitung bis ausreichende Information vorliegt)?
- Die Frage der Ressourcen: Welche Ressourcen (z.B. Arbeitsgedächtniskapazität, Aktivierung und Aktivierungszerfall, Zunahme von Interferenzen in Abhängigkeit der Informationsmenge) bestimmen die Beschränkung der Verarbeitungskapazität? Im Falle der Annahme von Parallelverarbeitung: Erfolgt die Verarbeitung mehrerer alternativer syntaktischer Strukturen beschränkt (je mehr parallel verfolgte Alternativen, desto größer die Auslastung der Ressourcen) oder unbeschränkt?

**Untersuchungsmethoden und -gegenstände** Um die Merkmale der menschlichen Fähigkeit syntaktischer Sprachverarbeitung zu erklären, wurden verschiedenste Modelle

für die zugrundeliegenden kognitiven Mechanismen und Strukturen vorgeschlagen. Um die Modelle empirisch zu validieren, haben sich Untersuchungsfelder als besonders fruchtbar herauskristallisiert, die zu empirisch beobachtbaren Problemen bei der Verarbeitung führen: die Verarbeitung von ambigen sowie von komplexen Sätzen. Beide Gegenstände – sowohl komplexe als auch ambige Sätze – werden in der Diskussion zu lokalen syntaktischen Kohärenzen wieder eine Rolle spielen. Die Grundidee ist, dass weniger komplexe oder (im Falle von Ambiguitäten) präferierte Strukturen schneller oder müheloser verarbeitet werden bzw. komplexere oder dispräferierte Strukturen entsprechend langsamer und mit höherer Fehleranfälligkeit.

Wodurch aber werden Sätze komplexer oder einfacher? Welcher Natur sind unsere Präferenzen und wodurch kommen sie zustande? Welche Information kommt dabei zu welchem Zeitpunkt zum Tragen? Um solche Fragen zu beantworten, wurde mit verschiedenen empirischen Methoden die Verarbeitung einer Reihe verschiedener Satzstrukturen und inhaltlicher Manipulationen untersucht, anhand derer die Eigenschaften und die Grenzen des menschlichen Sprachverarbeitungssystems kontrastiv identifiziert werden können (für eine Übersicht siehe z.B. Altmann, 1998, 2001; Hemforth & Konieczny, 2002; Treiman, Clifton Jr., Meyer & Wurm, 2003). Hierfür wurden Sätze mit lokaler oder globaler Ambiguität verwendet, d.h. mit Mehrdeutigkeit, die sich entweder im Verlauf des Satzes auflöst oder nach Satzende bestehen bleibt. Diese Eigenschaft weisen zum Beispiel Sätze mit reduzierten Relativsätzen, Sätze mit Präpositionalphrasen mit alternativen Anbindungsmöglichkeiten, Sätze mit direkt an das Matrixverb gebundenen Komplementsätzen, Sätze mit Komplementsätzen mit alternativer Nominalphrasen-Anbindung, oder Sätze mit Komplementsatz-/Relativsatz-Ambiguität auf. Weiterhin wurden Sätze mit erhöhter Komplexität verwendet, zum Beispiel Sätze mit unregelmäßiger und selten verwendeter Satzstruktur oder Sätze mit mehrfach rekursiver Einbettung (vgl. Bsp. 3 in Kap. 3.2.1). Außer der Ambiguität und der Komplexität wurden als Kernfaktoren auch Kontext-Effekte sowie Priming-Effekte untersucht. Im nachfolgenden Unterkapitel werden wir Sätze vorstellen, die ein strukturelles Phänomen enthalten, das erst in der jüngeren Forschung verstärkt Beachtung gefunden hat: Lokale syntaktische Kohärenzen. Ein Klassiker in der psycholinguistischen Forschung, der für dieses Phänomen relevant ist, sind die sogenannten Holzweg-Sätze (Garden Path Sentences, vgl. Frazier, 1987; Bever, 1970). In Holzweg-Sätzen entsteht Komplexität dadurch, dass der Leser den strukturellen Präferenzen entsprechend inkrementell eine Struktur des Satzes aufbaut. Für die gewählte Interpretation wird im Verlauf des Satzes durch später hinzukommende Information klar, dass sie falsch ist, weshalb die zuvor getroffene strukturelle Entscheidung zugunsten einer alternativen strukturellen Analyse revidiert werden muss.

Abhängig von der verwendeten Methode und dem gemessenen Verhalten finden sich mit den oben genannten Satztypen Effekte, die einen Hinweis auf die zugrundeliegenden kognitiven Prozesse, Strukturen und Interaktionen geben. So finden sich zum Beispiel in Lesezeitmessungen positionsbezogen erhöhte Latenzen. In Eyetracking-Studien finden sich Rücksprünge und vermehrtes Lesen von Positionen. Akzeptabilitätsratings ergeben bessere oder schlechtere Bewertungen eines Satzes. Verständnisfragen können mit unterschiedlicher Erfolgsquote richtig oder falsch beantwortet werden. In Primingstudien finden sich schnellere oder langsamere Reaktions- bzw. Entscheidungszeiten. Im experimentell kontrollierten Kontrast zeigt sich, welche syntaktischen Strukturen zu vermehrter Fehl-

interpretation oder vermehrtem Verarbeitungsaufwand führen, welche Informationen zu Verarbeitungsunterschieden führen und auf welchen Satzpositionen sich Interaktionen mit anderen Aufgaben finden. Aus solchen Effekten können Rückschlüsse gezogen werden, welche Eigenschaften der Sprache einen erhöhten kognitiven Aufwand erfordern und auf welchen kognitiven Beschränkungen, Mechanismen und Eigenschaften ihre Verarbeitung basiert. Die Identifikation der Auswirkung verschiedener Faktoren ermöglicht im Vergleich mit spezifischen Vorhersagen verschiedener Theorien und Modelle die Falsifizierung von alternativen Erklärungsansätzen. Ein nicht mehr ganz aktueller, aber sehr detailliert aufbereiteter Vergleich von Vorhersagen verschiedener Modelle der Satzverarbeitung und empirischen Befunden aus Studien mit verschiedenen psycholinguistischen Methoden findet sich bei Mitchell (1994).

**Ansätze und Modelle** In der Psycholinguistik wurden verschiedene Erklärungsmodelle entwickelt, um die Eigenschaften der Sprachverarbeitung auf Prinzipien und Mechanismen des Sprachverarbeitungssystems zurückzuführen und anhand dieser empirische Befunde für solche Sätze zu verstehen und vorherzusagen. Dabei können die Ansätze hinsichtlich verschiedener Aspekte und Grundannahmen unterschieden werden. Eine klare Abgrenzung verschiedener Ansätze ist allerdings schwierig, da Theorien mit sehr unterschiedlichen Grundlagen in einem Aspekt zum Teil auf gleiche Grundlagen in einem anderen Aspekt zugreifen. Vergleichende Übersichten über verschiedene theoretische Ansätze finden sich in der einschlägigen psycholinguistischen Literatur (vgl. u.a. Mitchell, 1994; Konieczny, 1996; Crocker et al., 2000; Vosse & Kempen, 2000; Altmann, 2001; Clifton & Duffy, 2001; Hemforth & Konieczny, 2002; Treiman et al., 2003; Altmann, 2006). Die folgende Aufführung gibt einen groben Überblick:

- Theoretischer Ausgangspunkt zur Erklärung der Verarbeitungseigenschaften: sprachliche Kompetenz, beispielsweise in grammatikbasierten Ansätzen in der Tradition der Generativen Grammatik, etwa der Derivational Theory of Complexity (DTC, z.B. Slobin, 1966). Performanz, zum Beispiel in strukturbasierten heuristischen Modellen (Bever, 1970; Frazier, 1987; Konieczny, Hemforth, Scheepers & Strube, 1997). Aufhebung der Kompetenz-Performanz-Dichotomie, etwa in erfahrungsbasierten (usage-based approaches, Bybee & McClelland, 2005), konnektionistischen (Elman et al., 1996; Elman, 2004) oder statistischen (Gennari & MacDonald, 2008; Wells et al., 2009) Ansätzen.
- Kognitive Ressourcen und ihre Beschränkungen sowie Ökonomiepostulate als Erklärungsgrundlage für empirisch beobachtbare Präferenzen, Komplexitätseffekte und Ambiguitätseffekte. Sparsamkeit beim Aufbau syntaktischer Strukturen (bei heuristik-basierten Ansätzen, etwa beim Garden-Path-Modell, Frazier, 1987). Verbales Arbeitsgedächtnis (als einzelne oder multiple Ressource) als Hauptquelle für Komplexität und individuelle Differenzen (z.B. Just & Carpenter, 1992; Waters & Caplan, 1996). Begrenzte Aktivierungsenergie und Aktivierungszerfall, zum Beispiel in der Dependency Locality Theory (DLT, Gibson, 1998, 2000). Sparsamkeit des für einen Satz aufgebauten mentalen Modells (Crain & Steedman, 1985). Hinsichtlich des Aspekts der Ressourcen stehen Ansätze serieller (Frazier, 1987) und beschränkt paralleler (Gibson, 2000) Verarbeitung Ansätzen unbegrenzter Parallelverarbeitung entgegen (Elman, 1991; Christiansen & Chater, 1999b). Unterschiede der Erfahrung

und der statistischen Lernfähigkeit (MacDonald & Christiansen, 2002; Gennari & MacDonald, 2008; Wells et al., 2009).

- Zentrale kognitive Prozesse und Mechanismen als Erklärungsgrundlage für empirisch beobachtbare Präferenzen sowie Komplexitäts- und Ambiguitätseffekte. Bewegungsoperationen für strukturelle Transformationen (DTC, Slobin, 1966). Heuristiken zur Wahrung von Ökonomieprinzipien (z.B. Minimal Attachment und Late Closure, Frazier, 1987). Reanalyse beim Scheitern der initialen Analyse (Ferreira & Henderson, 1991; Sturt, Pickering & Crocker, 1999). Mechanismen zum Abruf temporär enkodierter Informationen aus dem Arbeitsgedächtnis zur Auflösung von Abhängigkeiten, etwa in der DLT (Gibson, 2000) oder in Cue-Based Parsing-Ansätzen (R. L. Lewis, Vasishth & van Dyke, 2005; R. L. Lewis & Vasishth, 2006). Probabilistische Mechanismen und Constraint-Satisfaction-Prinzipien (McClelland et al., 1989; MacDonald, Pearlmutter & Seidenberg, 1994; MacDonald & Seidenberg, 2006). Antizipation von Folgestrukturen, etwa in konnektionistischen Modellen (Altmann, 1999; MacDonald & Christiansen, 2002; Konieczny & Döring, 2003) oder nach wahrscheinlichkeitstheoretischen Bemessungen in probabilistischen Modellen (Surprisal- und Entropy Reduction-Modelle, z.B. Hale, 2003; R. Levy, 2008).
- Zentrale Einheiten der mentalen Repräsentation und zentrale Einheiten für Parsingpräferenzen: Operationen oder Heuristiken basierend auf Phrasenstrukturbäumen (Frazier, 1987; Bever, 1970). Zustand oder Trajektorie im Repräsentationsraum (Elman, 1991; Tabor & Tanenhaus, 1999). Mentales Situationsmodell (Crain & Steedman, 1985) und diskursbasierte Strategien zur Eingliederung von Folgeelementen in den gegebenen Diskurskontext (Altmann & Steedman, 1988). Zentrale Bedeutung lexikalischer Einheiten und des mentalen Lexikons, etwa in der Lexical Functional Grammar (R. Kaplan & Bresnan, 1982) oder in Constraint-basierten Modellen (McRae, Spivey-Knowlton & Tanenhaus, 1998). Interaktion mit Merkmalen der Zielsprache wie Frequenz und Regularität (MacDonald & Christiansen, 2002; Gennari & MacDonald, 2008; Wells et al., 2009).
- Interaktion oder Modularität der Komponenten verschiedener Verarbeitungs- und Informationsaspekte. Modulare Komponenten mit Primat der syntaktischen Analyse (syntax-first-Modelle, z.B. Frazier, 1987). Interaktive Ansätze, etwa Multiple-Constraint-Ansätze (MacDonald et al., 1994; Tanenhaus et al., 1995; MacDonald & Seidenberg, 2006; Altmann & Mirkovic, 2009) und Ansätze zur Integration visueller und linguistischer Informationen (Tanenhaus et al., 1995; Altmann & Kamide, 1999; Mayberry et al., 2009).

**Aktuelle Phänomene und Debatten** Einige aktuelle Beobachtungen und Debatten der jüngeren Forschungsgeschichte haben die Überprüfung, Konkretisierung und Differenzierung verschiedener Modelle und Theorien zur menschlichen Sprachverarbeitung weiter vorangetrieben. Sie sind für die in den folgenden Unterkapiteln näher besprochenen Gegenstände von besonderer Relevanz. Zu diesen neueren Phänomenen zählen Lokalitätseffekte, Antilokalitätseffekte sowie Effekte lokaler syntaktischer Kohärenzen. Alle haben etwas damit zu tun, welche Aspekte der syntaktischen Struktur für Komplexität und Interferenzen bei der Anbindung lexikalischer Elemente in den syntaktischen Kontext und die Auflösung

von Abhängigkeiten verantwortlich sind. Lokalitätseffekte sind abhängig von der Distanz zwischen abhängigen Elementen und der Menge intervenierender bzw. konkurrierender Elemente zwischen diesen, wobei kürzere Distanzen zu einfacherer und akkuraterer Verarbeitung führen als große Distanzen (vgl. z.B. Gibson, 1998, 2000; van Dyke & Lewis, 2003; R. L. Lewis et al., 2005). Antilokalitätseffekte führen zum genau entgegengesetzten Ergebnis: Je weiter das abhängige Element vom Kopf entfernt ist, desto einfacher wird seine Verarbeitung. Dies könnte daran liegen, dass entsprechend mehr Information zur Antizipation des Elements zur Verfügung steht (vgl. z.B. Konieczny, 2000; Konieczny & Döring, 2003; T. F. Jaeger et al., submitted). Effekte lokaler syntaktischer Interferenz führen zur falschen Anbindung von Elementen, die nur in einem lokal begrenzten Fenster korrekt wären, im globalen Kontext aber ungrammatisch sind (vgl. z.B. Tabor et al., 2004; Konieczny, Müller, Hachmann et al., 2009; Konieczny & Müller, 2010).

Unter anderem in Bezug auf das letztgenannte Phänomen entwickelt sich aktuell eine Debatte, die sich mit der Rationalität des menschlichen Sprachverarbeitungssystems auseinandersetzt. Es wird dabei in Frage gestellt, ob hierbei tatsächlich (a) alle zur Verfügung stehenden Informationen für die Verarbeitung von Sätzen benutzt werden, (b) die globale Konsistenz der konstruierten Interpretation gewahrt bleibt, d.h. bei der Integration neuer Elemente der globale syntaktische Kontext berücksichtigt wird, und (c) nur global konsistente Interpretationen in Erwägung gezogen werden, während inkonsistente Fortsetzungen oder unnötige Restrukturierungen der bisher aufgebauten Struktur unmittelbar ausgeschlossen werden. In den meisten Ansätzen wurde bis dato axiomatisch angenommen, dass das Sprachverarbeitungssystem rational und unter der Wahrung globaler Konsistenz arbeitet. In einigen aktuell entwickelten Ansätzen wird versucht, die Annahme der Rationalität aufrechtzuerhalten und scheinbar irrationale Phänomene wie etwa lokale syntaktische Kohärenzen auf rationaler Ebene über Interferenzen, Interaktionen statistischer Eigenschaften und Ressourcenbeschränkungen zu erklären (z.B. Gibson, 2006; van Dyke, 2007; Bicknell & Levy, 2009; Hale, 2011). In anderen Theorien und Modellen werden Ansätze zum Sprachverarbeitungssystem entwickelt, in denen die Annahme der Rationalität und der globalen Konsistenz der Interpretation irrelevant ist oder abgelehnt wird, darunter dynamische und konnektionistische Modelle (u.a. Tabor & Tanenhaus, 1999; Christiansen & Chater, 1999b; Konieczny & Müller, 2010) und der Ansatz des Good Enough-Parsing (Ferreira et al., 2002; Ferreira & Patson, 2007). Dem letztgenannten Ansatz zufolge ist die Verarbeitung nicht auf Perfektion und Vollständigkeit ausgelegt, sondern darauf, gut genug und dadurch schnell und robust zu sein – was in den meisten Fällen für die erfolgreiche Verarbeitung genügt.

Für die Verarbeitung von Anaphern – d.h. die Auflösung ihres koreferenziellen Bezugs – werden mitunter Fragen relevant, die bereits seit längerem in der Forschung diskutiert werden: so die Frage, zu welchem Zeitpunkt verschiedener Aspekte eines Satzes verarbeitet werden und inwiefern die zuständigen Prozesse modular oder interaktiv erfolgen. Dabei steht mitunter im Vordergrund, ob Präferenzen und Beschränkungen der Auswahl möglicher und zwingender Koreferenten durch strukturell-syntaktische, semantische, pragmatische oder diskursive Bedingungen bestimmt werden, und welche Bedingung mit welcher Gewichtung und zu welchem Zeitpunkt auf die Auswahl Einfluss nimmt (vgl. u.a. Kaiser et al., 2009; Runner et al., 2006; van Gompel & Majid, 2004; Sturt, 2003; Badecker & Straub, 2002).

## 5.2. Lokale syntaktische Kohärenzen

Lokale syntaktische Kohärenzen (LSK) sind Subsequenzen innerhalb von Sätzen, die isoliert betrachtet einen grammatikalisch korrekten Satz formen. Im globalen Kontext des Gesamtsatzes jedoch kann die Interpretation dieser Subsequenz als abgeschlossene Einheit nicht aufrechterhalten werden, da die Bestandteile an andere Teile des Gesamtsatzes gebunden werden müssen. D.h. eine Interpretation der Subsequenz als geschlossene syntaktische Einheit führt zu einer Verletzung des globalen syntaktischen Kontextes. Bei sogenannten Gardenpath-Sätzen (Frazier, 1987) kann die Verletzung der korrekten Satzstruktur erst nach der kritischen Position erkannt werden – bis zu diesem Auflösungspunkt ist die getroffene Wahl der Struktur möglich (man wird bei entsprechenden Anbindungspräferenzen also dazu verleitet, die falsche Strukturwahl zu treffen). Demgegenüber steht bei LSK schon zu Beginn der kritischen Position fest, dass deren Satzstruktur nicht mit der bis dahin inkrementell aufgebauten Satzstruktur übereinstimmt. Folglich sollte es ausgeschlossen sein, dass diese Interpretationsmöglichkeit in Betracht gezogen wird.<sup>2</sup>

Um dies an einem Beispiel zu illustrieren (aus Konieczny, 2005): Beispielsatz 5 enthält eine LSK. Diese entspricht isoliert betrachtet einem grammatikalischen Satz des Deutschen (Bsp. 5a): Sie kann als einfacher Hauptsatz mit Subjekt, transitivem Verb, Akkusativ-Objekt und Präpositionalphrase gelesen werden. Diese Struktur kann aber im globalen Satzgefüge nicht aufrechterhalten werden (Bsp. 5b). Die Satzsequenz steht in einem Nebensatz, eingeleitet durch den Komplementierer *dass*. Folglich muss das finite Verb dieses Teilsatzes nach den Regeln der Deutschen Grammatik am Ende desselben stehen, und folglich kann es sich bei dem Wort *überrascht* nur um ein Adverb handeln, nicht jedoch um das finite Verb des Teilsatzes. Also sollte die LSK in Anbetracht der inkrementell aufgebauten globalen Struktur des Satzes nicht als mögliche interpretierbare Einheit in Betracht gezogen werden.

- (5) Die Nachricht, dass die Astronautin überrascht den Außerirdischen vom Mars entdeckte, machte bald die Runde.
- a. Die Astronautin überrascht den Außerirdischen vom Mars.
  - b. Die Nachricht, dass *die Astronautin überrascht den Außerirdischen vom Mars* entdeckte, machte bald die Runde.

Der Begriff lokale Kohärenz darf dabei nicht mit dem diskurstheoretischen Begriff der lokalen Kohärenz verwechselt werden. Letzterer ist ein zentraler Faktor der Herstellung globaler Kohärenz über einen größeren Text hinweg: Er beschreibt den textuellen Zusammenhang beim Übergang zwischen verschiedenen Sätzen. Regularitäten hierfür wurden

<sup>2</sup>Lokale syntaktische Kohärenzen können also inkongruente Fortsetzungen innerhalb des Satzes sein. Die lokale Sequenz kann eigentlich nicht als Fortsetzung des bislang gelesenen Satzes in Betracht gezogen werden, da sonst die grammatikalische Kongruenz dieses Satzes verletzt werden müsste. Die Interpretation der lokalen Sequenz als selbständige in sich kohärente Einheit verlangt also die Verletzung des bis dorthin verarbeiteten Satzes und müsste folglich von vornherein von den Interpretationsmöglichkeiten ausgeschlossen werden. In den Worten von Hale (2011, S. 22): „On the face of it, local coherence seems different from garden pathing. In classic temporary ambiguities, some analysis of the ambiguous region turns out to be grammatically inconsistent with material that eventually arrives on its right. By contrast, a locally coherent substring may be inconsistent as well with material on its left.“

zum Beispiel von Grosz, Joshi und Weinstein (1995) beschrieben. Zur Abgrenzung von dieser Verwendung sprechen wir im Folgenden immer von lokalen *syntaktischen* Kohärenzen bzw. verwenden die eingeführte Abkürzung LSK.

### 5.2.1. Einfluss auf die Satzverarbeitung: Interferenz

Dass solche lokal kohärenten Substrukturen aber de facto mit der globalen Interpretation des Satzes interferieren können, wurde zuerst von Tabor et al. (2004) gezeigt. In einer Serie von Leseexperimenten mit Self-Paced Reading<sup>3</sup> im Moving-Window-Paradigma<sup>4</sup> untersuchen sie den Einfluss von LSK auf die Lesezeit auf bestimmten Positionen. In ihren Experimental-Stimuli verwenden sie Sätze, die eine syntaktisch kohärente Subsequenz mit Hauptsatzstruktur enthalten (Beispiel 6a). Diese kontrastierten sie mit Sätzen, die eine kohärente Lesart der Subsequenz nicht zulassen (6b, 6c, 6d). Das Material wurde nach den Faktoren Relativsatztyp (reduzierter vs. nicht reduzierter Relativsatz) und Verbambiguität (ambig vs. nicht ambig) in einem 2×2-Design in vier Bedingungen kontrastiert.

- (6) The coach smiled at...
- a. the player tossed a frisbee (reduziert, ambig)
  - b. the player who was tossed a frisbee (nicht reduziert, ambig)
  - c. the player thrown a frisbee (reduziert, nicht ambig)
  - d. the player who was thrown a frisbee (nicht reduziert, nicht ambig)
- ...by the opposing team.

Das Verb *tossed* ist morphologisch ambig – es kann isoliert betrachtet gleichermaßen im Partizip Perfekt stehen oder aber als finites Hauptverb in der einfachen Vergangenheitsform.

---

<sup>3</sup>Beim Self-Paced Reading (Just, Carpenter & Woolley, 1982) wird den VTn Leseinheit für Leseinheit (z.B. Wort für Wort, wobei auch kleinere oder größere Einheiten wie Phrasen möglich sind) auf dem Computerbildschirm dargeboten, wobei die Darbietung durch Tastendruck auf einem Eingabegerät erfolgt. Die VTn bestimmen also selbst die Geschwindigkeit der Darbietung des neuen Materials. Aufgezeichnet wird die Zeit zwischen zwei Tastendrücker. Aus dieser Latenz wird der Zeitaufwand für das entsprechende Wort abgeleitet. Diese Methode hat den Nachteil, dass außer der eigentlichen Aufgabe (dem Lesen und Verstehen des Satzes) der Tastendruck als zweite motorische Aufgabe ausgeführt werden muss, die ebenfalls kognitive Ressourcen erfordert.

<sup>4</sup>Als Darbietungsformen sind Moving Window und Stationary Window zu unterscheiden. Bei der Variante Stationary Window werden die Wörter immer an derselben Stelle eingeblendet. Bei der Variante Moving Window werden die Wörter sukzessive rechts angefügt, so dass ein normaler Lesefluss mit Blickbewegung in Leserichtung erhalten bleibt. Für die Variante Moving Window gibt es eine weitere Unterscheidung für das Vorgehen vor und nach Darbietung der Wörter: In der unmaskierten Variante wird jedes Wort an der jeweils nächsten Stelle eingeblendet. In der maskierten Variante hingegen steht bereits für den kompletten Satz oder Text ein Stellvertreter auf dem Bildschirm, wobei alle Buchstaben des Satzes maskiert werden. Sukzessive wird nun auf allen Stellvertreterpositionen das jeweilige Wort eingeblendet. Diese Methode hat den Vorteil, dass eine natürlichere Blickbewegungssteuerung erfolgen kann und vorab Informationen über den kommenden Text verarbeitet werden können (z.B. Länge des nächsten Wortes). Eine weitere Variante betrifft die Permanenz der Wörter: Bleiben diese nach dem Erscheinen stehen (kumulativ) oder werden sie wieder ausgeblendet bzw. maskiert (nicht-kumulativ). Zwar ist es natürlicher, wenn der Text nach dem ersten Lesen stehen bleibt und Rücksprünge zu vormals gelesenen Text ermöglicht werden. Diese Rücksprünge entgehen aber natürlich der Aufzeichnung, und verzerren die Daten für die Lesezeiten. Zudem begünstigt die kumulative Variante ein strategisches Vorgehen.

Deshalb enthält der Satz 6a eine LSK: Mit der Hauptverb-Lesart im reduzierten Relativsatz kann die Subsequenz *The player tossed a frisbee* gebildet werden, die isoliert betrachtet einen kohärenten und vollständigen Hauptsatz bildet. Alle anderen Bedingungen hingegen enthalten keine LSK. In den Sätzen 6b und 6d wird der Relativsatz nicht reduziert und folglich mit *who was* eingeleitet. Die Sequenz bildet damit keinen selbständigen kohärenten Satz mehr. In Satz 6c hingegen wird *tossed* durch das Verb *thrown* ersetzt, das durch die unregelmäßige Flexionsmorphologie eindeutig als Partizip Perfekt erkennbar ist und damit die Hauptsatzlesart der Subsequenz ebenfalls unterbindet.

Wie bereits betont, ist die Interpretation der Subsequenz in 6a inkonsistent mit der Interpretation der globalen Satzstruktur. Ihre Verarbeitung würde eine Verletzung der grammatikalischen Kongruenz der Sequenz mit der vorhergegangenen Satzsequenz erfordern: *the player* ist zwingend Objekt des vorausgegangenen Teilsatzes *The coach smiled at the player..* und kann folglich nicht das Subjekt eines neuen Satzes sein. Dementsprechend müsste die Interpretation der LSK bei der Verarbeitung des Satzes unmittelbar ausgeschlossen werden.

Tabor et al. (2004) führen insgesamt drei Self-Paced-Reading-Experimente mit jeweils anschließenden Verständnisfragen bzw. im dritten Experiment mit einem Rating der Grammatikalität der Sätze durch. Wie sie im ersten Experiment zeigen, interferieren LSK entgegen den Erwartungen einer grammatikalisch konsistenten Interpretation tatsächlich mit der globalen Interpretation und verletzen damit die grammatische Kongruenz des inkrementell aufgebauten syntaktischen Kontextes. In der kritischen Satzregion, die das Relativsatzverb und die folgenden drei Wörter enthält, zeigen sie einen Haupteffekt für den Faktor Relativsatztyp mit höheren Lesezeiten für reduzierte Relativsätze (6a und 6c), sowie eine Interaktion zwischen den Faktoren Relativsatztyp und Verbambiguität. In dieser Interaktion findet sich ein Unterschied für Ambiguität mit höheren Lesezeiten für das ambige Verb nur für die Bedingungen mit reduziertem Relativsatz (6a vs. 6c), nicht aber in den nicht-reduzierten Versionen. Am deutlichsten werden die Effekte auf der Wortposition unmittelbar nach dem Relativsatzverb.

In Experiment 2 zeigen die Autoren des Weiteren, dass dieser Effekt nur dann zustande kommt, wenn die LSK auch semantisch kohärent ist. Hierfür wurde in den verwendeten Stimuli nicht das Relativsatzverb variiert, sondern das Objekt, auf das sich der Relativsatz bezieht und das in der LSK das potentielle Subjekt stellt. Dieses Objekt ist entweder belebt oder unbelebt, und kann in der LSK nur im ersteren Fall (Beispiel 7a) Agent des Relativsatzverbs und damit Satzsubjekt der LSK sein. Als Beispiel hierzu: Nur ein Gefangener kann etwas transportieren (*The prisoner<sub>Subj</sub> transported...*), nicht aber Gold (*\*The gold<sub>Subj</sub> transported...*). Das Material wurde dementsprechend nach den Faktoren Relativsatztyp (reduzierter vs. nicht reduzierter Relativsatz) und Objektbelebtheit (animiert vs. nicht animiert) in einem 2×2-Design in vier Bedingungen kontrastiert. Auch in diesem Experiment zeigen Tabor und Kollegen in der kritischen Satzregion (das Relativsatzverb und die folgenden drei Wörter) einen Haupteffekt für den Faktor Reduziertheit. Zudem findet sich hier, wie von den Autoren erwartet, eine Interaktion zwischen Reduziertheit und Belebtheit: Nur für die semantisch konsistente Bedingung ergeben sich erhöhte Lesezeiten entsprechend den Befunden des ersten Experiments.

- (7) The bandit worried about...
- a. the prisoner transported the whole way (reduziert, animiert)
  - b. the prisoner who was transported the whole way (nicht reduziert, animiert)
  - c. the gold transported the whole way (reduziert, nicht animiert)
  - d. the gold that was transported the whole way (nicht reduziert, nicht animiert)
- ...by the capricious guards.

Im dritten Experiment wurden die Sätze aus den beiden ersten Experimenten verwendet und in gekürzten Versionen präsentiert. Über ein jeweils direkt anschließendes Rating der Grammatikalität mit binärem ja-/nein-Rating sollte festgestellt werden, ob sich Unterschiede für die grammatische Akzeptabilität der Sätze finden. Unterschiede könnten aus den Differenzen in der kritischen Satzregion resultieren, auf der sich die Effekte der LSK am deutlichsten zeigen. Zugleich sollte damit kontrolliert werden, ob die Effekte möglicherweise durch ein generell gestörtes Parsing verursacht werden. So könnte die Extraktion von Rezipienten (in Experiment 1) bzw. unbelebten Objekten (in Experiment 2) in Relativsätzen eher ungewöhnlich sein. Sollte dies der Fall sein, so sollten sich allgemein schlechte Bewertungen für die Sätze zeigen. Tabor und Kollegen demonstrieren jedoch dieselben Effektmuster wie in den vorhergehenden Experimenten – nur in einem anderen experimentellen Maß.

Tabor et al. (2004) interpretieren die Ergebnisse der Studie als Evidenz dafür, dass lokal (semantisch und syntaktisch) kohärente Subsequenzen mit der globalen Interpretation interferieren, auch wenn sie global inkonsistent sind und im inkrementellen Leseprozess eigentlich sofort ausgeschlossen sein müssten. Da LSK mit der globalen Interpretation interferieren, führen sie zu erhöhten Lesezeiten in der lokalen Kohärenz.

### 5.2.2. Einfluss auf die Satzverarbeitung: Interpretation

Die Forschungsgruppe um Konieczny (u.a. Konieczny, 2005; Konieczny & Müller, 2007; Konieczny, Müller, Hachmann et al., 2009) demonstriert Ergebnisse für das Deutsche, die den Befund von Tabor et al. (2004) stützen, dabei aber seine Tragweite wesentlich ausdehnen. Ihrer Hypothese nach interferieren LSK nicht nur mit der globalen Interpretation – vielmehr werden sie auch selbst Gegenstand der Interpretation. In den Studien kamen vor allem die folgenden psycholinguistischen Methoden zum Einsatz: Blickbewegungsstudien im Visual-World-Paradigma, Blickbewegungsstudien beim Lesen und Anomaly-Detection-Studien. Durch die Verwendung des Visual-World-Paradigma konnte zudem untersucht werden, ob die Effekte von Tabor und Kollegen nur auf das Lesen von Sätzen beschränkt ist oder auch beim Verstehen gesprochener Sprache auftreten. Entsprechende Befunde unterstützen die Annahme der Allgemeinheit des Effekts für das ganze Sprachverarbeitungssystem.

**Anomaly Detection** Die früheste Experimentalserie mit Anomaly-Detection-Aufgabe<sup>5</sup> wurde durch Befunde aus konnektionistischen Simulationen inspiriert. In ihr wird der

---

<sup>5</sup>Bei Anomaly Detection (vgl. Boland, Tanenhaus, Carlson & Garnsey, 1989; Konieczny, 2005) müssen die VTn per Tastendruck so schnell wie möglich signalisieren, wenn ein gegebener Satz ihrer Meinung

Einfluss von sogenannten False Alarms als Quelle von LSK-Effekten untersucht (vgl. Kap. 3.2.3). Die Sätze, die von Konieczny (2005) in diesem Paradigma verwendet werden, beinhalten jeweils einen zentral eingebetteten Relativsatz mit ungrammatischer Fortsetzung (vgl. Bsp. 8). Sie variieren dabei in einem  $2 \times 2$ -Design in den Faktoren Satztyp des Relativsatzes (Subjekt- vs. Objektextraktion) und Kohärenz der ungrammatischen Fortsetzung (LSK-Erweiterung vs. inkohärente Erweiterung).

- (8) a. Der Abgeordnete, der den Journalisten beschimpft \*[den Politiker], liefert die Beweise. (Subjekt-RS, inkohärent)  
 b. Der Abgeordnete, der den Journalisten beschimpft \*[der Politiker], liefert die Beweise. (Subjekt-RS, kohärent)  
 c. Der Abgeordnete, den der Journalist beschimpft \*[der Politiker], lieferte die Beweise. (Objekt-RS, inkohärent)  
 d. Der Abgeordnete, den der Journalist beschimpft \*[den Politiker], liefert die Beweise. (Objekt-RS, kohärent)

Entsprechend der Befunde aus den falschen Aktivierungen lokal kohärenter Folgewörter aus den SRN-Simulationen (Konieczny & Ruh, 2003; Konieczny, Ruh & Müller, 2009) und der Annahme der Verarbeitung von LSK wurde erwartet, dass die ungrammatischen (im Beispiel mit Asterisk und eckigen Klammern markierten), aber lokal kohärenten Fortsetzungen – *Den Journalisten beschimpft der Politiker.* in Bsp. 8b und *Der Journalist beschimpft den Politiker.* in Bsp. 8d – langsamer bzw. später erkannt werden als ihre lokal inkohärenten Gegenbeispiele. Denn zumindest in der initialen lokalen Verarbeitung müssten sie der False-Alarms-Hypothese zufolge ebenfalls aktiviert werden. Dies sollte zu längeren Reaktionszeiten in der Anomaly-Detection-Aufgabe führen. Zudem wurde angenommen, dass der Effekt für Objekt-Relativsätze stärker ausfallen müsste, da die darin enthaltene LSK eine regelmäßige Wortstellung (Subjekt-Verb-Objekt) aufweist.

Tatsächlich findet Konieczny (2005) längere Reaktionszeiten für die LSK-Bedingung im Falle der Subjekt-Relativsätze: ein Haupteffekt für den Faktor Kohärenz und eine Interaktion zwischen beiden Faktoren. Dies ist ein Hinweis auf die Verarbeitung von LSK. Die Tatsache, dass der Effekt von LSK nur für die Sätze mit Subjekt-Extraktion gefunden wird, bietet zwei mögliche Erklärungen: (a) Die Verarbeitung ist durch Ressourcen, etwa die Kapazität des verbalen Arbeitsgedächtnisses, determiniert. Nur Subjekt-Relativsätze beanspruchen diese Ressource in solch geringem Maße, dass auch der lokale Parse parallel verfolgt werden kann. (b) Die Verarbeitung ist durch Erfahrung determiniert, wobei verschiedene lokale Kohärenzen um Aktivierung konkurrieren. In beiden Satzbedingungen findet sich die im Deutschen häufig erscheinende Sequenz Verb-Subjekt (z.B. ... *beschimpft der Politiker.*), die beispielsweise in Fragesätzen und in Sätzen mit Topikalisierung etwa des Objekts oder eines Adverbs zu finden ist. In Objekt-Relativsätzen widersprechen sich

---

nach semantisch oder syntaktisch nicht mehr sinnvoll oder korrekt ist. In der Regel werden die Sätze dazu mit der Methode Rapid Serial Visual Presentation (RSVP) dargeboten, bei der die Leseinheiten im Unterschied zum Self-Paced Reading mit computergesteuerter Taktung dargeboten werden. Die VTn haben also keinen Einfluss auf die Darbietungszeit. Anomaly Detection kann als eine Art implizites Grammatikalitätsrating verstanden werden: Den VTn werden in vorgegebener Geschwindigkeit Wort für Wort Sätze präsentiert, zum Beispiel visuell auf einem Bildschirm. Ihre Aufgabe ist es, eine Taste zu drücken, sobald der Satz sinnlos oder ungrammatisch wird.

die Vorhersagen für diese kleinere und die größere LSK, während sie sich in Subjekt-Relativsätzen überschneiden, und damit zum gefundenen Effekt führen.

Ein Problem der Anomaly-Detection-Experimente hinsichtlich ihrer Interpretation und ihrer Aussage für natürliche Sprachverarbeitung ist, dass die Stimuli des Experiments keine grammatikalischen Sätze sind. Im Zentrum steht damit eher die Untersuchung der Fehlererkennung in Sätzen. Kann darüber aber wirklich eine Aussage über die natürliche Verarbeitung von grammatikalischen Sätzen getroffen werden? Und können damit Eigenschaften des Einflusses von LSK auf die natürliche Satzverarbeitung identifiziert werden? Zumindest bieten die Ergebnisse der Anomaly Detection-Experimente einen Hinweis auf mögliche Effekte. Um aber weitergehende Fragen zu beantworten, wird der Einsatz anderer Methoden erforderlich, die das Messen entsprechender Effekte bei der Verarbeitung korrekter Sätze ermöglichen. Eyetracking-Experimente im Visual-World-Paradigma bieten diese Möglichkeit. Sie ermöglichen auch die Untersuchung der Frage, ob LSK nur Störeffekte sind, die mit der globalen Interpretation interferieren, oder ob sie tatsächlich selbst Gegenstand der Interpretation werden.

**Eyetracking im Visual-World-Paradigma** In einer Visual-World-Studie<sup>6</sup> wurden Sätze mit zentral eingebettetem Komplementsatz (eingeleitet mit *dass*) verwendet (Konieczny & Müller, 2007; Konieczny, Müller, Hachmann et al., 2009). Der Komplementsatz enthält an der Stelle direkt nach dem Satzsubjekt ein Partizip (globale Interpretation), dessen morphologische Form isoliert betrachtet ambig ist und (lokal) ein finites Hauptverb sein könnte. Diese Ambiguität ermöglicht eine ungrammatische Fortsetzung des Satzes, die innerhalb der lokalen Subsequenz grammatisch kohärent ist. Wir kennen das Beispiel bereits aus der Einleitung des Unterkapitels (vgl. Bsp. 5). Es handelt sich dabei um einen zentral eingebetteten Komplementsatz (vgl. Bsp. 9a), der bei Verwendung eines ambigen Partizips eine LSK enthält (*Die Astronautin überrascht den Außerirdischen vom Mars.*).

Die Sätze wurden in einem 2×2-Design in vier Bedingungen manipuliert: Die Zielbedingung (Satz 9a), die durch die lexikalisch ambige Form des Partizips, das auch als finites Hauptverb gelesen werden kann, eine LSK mit der Struktur eines regelmäßigen Hauptsatzes mit folgender Präpositionalphrase enthält, sowie drei Kontrollbedingungen. Diese wurden dadurch erstellt, dass in Bedingung 1 (Bsp. 9b) ein nicht-ambiges Adverb eingefügt wurde, sowie in den Bedingungen 3 (Bsp. 9c) und 4 (Bsp. 9d) ein zusätzliches nicht-ambiges Adverb (*gerade*), das vor das Adverb gestellt wurde, wodurch die lokale

---

<sup>6</sup>Das Visual-World-Paradigma (Cooper, 1974; Tanenhaus et al., 1995) ist eine Variante der Blickbewegungsmessung und funktioniert als modalitätsübergreifendes Verfahren. Hierbei werden den VTn auf dem Bildschirm visuelle Szenen präsentiert – dies können einzelne Referenten oder szenische Darstellungen mit beliebigem Komplexitätsgrad sein. Zugleich werden über Kopfhörer oder Lautsprecher Sätze dargeboten. Mit dem Eyetracking-System werden Fixationen und Sakkaden über die Abbildungen aufgezeichnet und in zeitlicher Auflösung parallel zum Verlauf des Satzes mit diesen assoziiert. Die Fixationen erlauben Rückschlüsse über die inkrementelle Interpretation des Satzes in hoher zeitlicher Auflösung. Befunde weisen auf eine starke zeitliche Verschränkung von dargebotenen Wörtern im Satz und Fixationen auf referierte Objekte hin, und unterstützen die Annahme sehr früher Integrationsprozesse visueller und linguistischer Informationen bei der Disambiguierung strukturell ambiger Sätze (Eberhard, Spivey-Knowlton, Sedivy & Tanenhaus, 1995; Tanenhaus et al., 1995). Der Analyse liegen kategoriale Daten des Vergleichs der Fixationswahrscheinlichkeit auf verschiedenen Zielregionen sowie Blickbewegungstrajektorien zwischen Zielregionen zugrunde. Eine ausführliche und aktuelle Beschreibung dieses Paradigmas findet sich bei Tanenhaus und Trueswell (2006).

Hauptsatzstruktur durchbrochen und die LSK zerstört wird. Die Adverbien der Kontrollbedingungen führen zu keiner substanziellen Veränderung der Interpretation des Nebensatzes. Die lokal isolierte Interpretation der LSK weicht jedoch – sollte sie bis auf die Ebene der semantischen Interpretation verarbeitet werden – von der globalen Interpretation des Nebensatzes ab.

- (9) Hier sieht man eine Astronautin und einen Außerirdischen in drei verschiedenen Szenen.
- a. Die Nachricht, dass die Astronautin überrascht den Außerirdischen vom Mars entdeckte, machte bald die Runde. (ambig, LSK)
  - b. Die Nachricht, dass die Astronautin ungläubig den Außerirdischen vom Mars entdeckte, machte bald die Runde. (nicht ambig, LSK)
  - c. Die Nachricht, dass die Astronautin gerade überrascht den Außerirdischen vom Mars entdeckte, machte bald die Runde. (ambig, keine LSK)
  - d. Die Nachricht, dass die Astronautin gerade ungläubig den Außerirdischen vom Mars entdeckte, machte bald die Runde. (nicht ambig, keine LSK)

Im Visual-World-Paradigma werden die Sätze auditiv präsentiert. Gleichzeitig zum Hören und Verstehen des Satzes müssen die VTn auf einem Bildschirm visuelle Darstellungen (Szenen, Objekte, etc.) betrachten. Im Zentrum des Experiments steht die Messung der Blickbewegungen in Abhängigkeit der aktuellen sequenziellen Positionen des auditiven Stimulus. In diesem Experiment wurden immer drei szenische Abbildungen gezeigt: (a) Eine Szene stellt die benannten Akteure des Satzes in einer Situation dar, die der Interpretation der LSK entspricht, (b) eine Szene entspricht der globalen Interpretation des Nebensatzes, (c) und eine dritte Szene zeigt die Akteure in einer Situation, die diesen Interpretationen neutral gegenübersteht. Im Experiment hörten die VTn jeweils einen Einleitungssatz, der die Akteure benennt (Bsp. 9), anschließend einen der entsprechenden Experimentalsätze.

In den aufgezeichneten Fixationen finden sich für Satzbedingung 1 mit LSK (Satz 9a) höhere Fixationsanteile auf der zur LSK korrespondierenden Szene als für Satzbedingung 2. Der Fixationsunterschied zeigt sich ab etwa 100 ms nach Off-Set des Partizips bzw. des Adverbs und dauert bis etwa 1700 ms an. Ebenfalls wurde eine erhöhte Fixationswahrscheinlichkeit auf der LSK-Szene in Bedingung 3 (Satz 9c) gegenüber Bedingung 4 festgestellt. Da in dieser LSK erst später ein Akteur benannt wird, findet sich der Effekt in einem späteren Zeitfenster zwischen 1500 ms und 2500 ms. Konieczny, Müller, Hachmann et al. (2009) sehen mit diesen Ergebnissen ihre Hypothese klar bestätigt, dass LSK nicht nur auf Ebene der syntaktischen Analyse mit der globalen Interpretation interferieren, sondern auch selbst zum Gegenstand der Interpretation werden.

Der Effekt von LSK wurde mehrfach in Studien mit alternativen syntaktischen Konstruktionen und anderen Kontrollbedingungen repliziert. Dabei wurden viele weitere Aspekte und Details des Effekts untersucht, mitunter auch in weiteren Experimentalparadigmen. Beispielsweise legen Konieczny, Müller, Baumann, Hachmann und Wolfer (2009) konvergierende Ergebnisse aus zwei Visual-World-Studien dar. Im ersten Experiment enthielt die Zielsatzkonstruktion einen Dativ-Relativsatz, wobei die LSK identisch wie im zuvor erläuterten Experiment ausfällt (vgl. Bsp. 10). Mit dieser Konstruktion können zwei kritische Aspekte des vorausgegangenen Experiments kontrolliert werden. Erstens findet sich in

Bedingung 3 (Bsp. 9c) eine weitere mögliche LSK, die die Struktur eines Hauptsatzes mit topikalisiertem Adverb aufweist (*Gerade überrascht den Außerirdischen...*). Zweitens könnten die Effekte als Holzweg-Effekte erklärt werden, denn die Sätze mit ambigem Partizip sind potenzielle Holzwegsätze (*Die Tatsache, dass die Astronautin gerade überrascht...*). In der Dativ-Konstruktion sind diese Möglichkeiten ausgeschlossen.

Im zweiten Experiment wurde untersucht, ob die Interpretation der LSK auch dann stattfindet, wenn die lokale Sequenz Satzgrenzen überschreitet. Hierzu verwenden die Autoren eine Konstruktion, die einen Objekt-Relativsatz enthält, der in einen vorangestellten Nebensatz eingebettet ist (vgl. Bsp. 11). Die LSK umfasst das Subjekt und das Verb des Relativsatzes sowie das Objekt des Matrixsatzes (*Die Floristin überraschte den Müllmann.*) und überschreitet damit die Relativsatzgrenze. Für diese Konstruktion ergibt sich der konvergierende Befund einer höheren Fixationswahrscheinlichkeit auf der Abbildung der lokal kohärenten Szene in der LSK-Bedingung. Dies kann als Hinweis darauf gesehen werden, dass LSK beim Verstehen gesprochener Sätze auch über Relativsatzgrenzen hinweg interpretiert werden. Allerdings setzen diese Ergebnisse voraus, dass die Sätze auf eine Weise bearbeitet werden, die den Einfluss der prosodischen Information auf die syntaktische Verarbeitung (Pausen, Betonung, Intonation) kontrolliert bzw. unterbindet.

- (10) Helmut, dem die Astronautin überrascht den Außerirdischen vom Mars zeigte, wollte seinen Augen nicht trauen.
- (11) Als der Postbote, den die Floristin überraschte, den Müllmann auf dem Marktplatz betrachtete, war es bereits dunkel.

**Eyetracking während des Lesens** In einem Experiment mit Blickbewegungsaufzeichnung während des Lesens<sup>7</sup> untersuchen Müller et al. (2009) den Einfluss semantischer Manipulationen innerhalb der LSK auf die Verarbeitung. Außerdem wurde dabei der Frage nachgegangen, ob innerhalb der LSK höhere Lesezeiten bzw. eine Erschwerung der Verarbeitung zu erwarten sind, wie in Ansätzen erwartet wird, die eine kompetitive Parallelverarbeitung annehmen, oder niedrigere Lesezeiten bzw. eine Erleichterung der Verarbeitung, wie sie in Ansätzen erwartet wird, die von Antizipation oder Projektion bei unbeschränkten Ressourcen ausgehen.

---

<sup>7</sup>Blickbewegungsaufzeichnung während des Lesens (Eyetracking while Reading) ist eine Methode, mit der die Probleme der Darbietungsgeschwindigkeit oder der Zweitaufgabe sowie der mangelnden Kontrolle von Rücksprüngen umgangen werden können. Großer Vorteil der Methode ist zum einen, dass sie wesentlich besser dem natürlichen Lesen entspricht, zum anderen werden unwillkürliche Handlungen aufgezeichnet (der VTn ist sich der aufgezeichneten Handlung also nicht bewusst). Der Text wird normalerweise vollständig auf einem Computerbildschirm dargeboten, wobei natürlich auch hier andere Möglichkeiten der Darbietung eingesetzt werden können. Zeitgleich werden die Fixationen, Fixationszeiten und Sakkaden gemessen (vgl. Rayner, 1998). Zur Aufzeichnung der Fixationen und Sakkaden werden verschiedene technische Verfahren verwendet. Die meisten modernen Verfahren verwenden dabei Videoaufzeichnungen der Augen, wobei mit einer Lichtquelle (Infrarot oder infrarotnahe Spektren) eine Hornhautreflektion erzeugt wird, über die die genaue Blickposition bestimmt werden kann. Über die Zeit, die auf ein Wort oder eine Region in verschiedenen Durchgängen geschaut wurde, kann auf den Verarbeitungsaufwand für das Wort geschlossen werden. Allerdings ist die Auswertung der Daten vergleichsweise komplex, unter anderem sind die Rohdaten um einen größeren Anteil an Stördaten zu bereinigen. Daneben sind verschiedene abhängige Maße zu unterscheiden, die im Gegenzug ein sehr hochaufgelöstes Bild der Online-Prozesse liefern.

Generell werden bei Experimenten mit Blickbewegungsaufzeichnung während des Lesens verschiedene abhängige Maße verglichen (vgl. z.B. Diskussionen in Konieczny, 1996; Rayner, 1998). So unterscheidet man unter anderem First Fixations und First Pass Reading Times, die Regression Path Duration, Second Pass Reading Times sowie die Total Reading Time. Für die First Fixations wird nur die jeweils erste Fixation auf einer Region  $n$  berücksichtigt. Mit den First Pass Reading Times wird die Summe der Fixationen auf einer Region  $n$  berechnet, bevor ein Blicksprung in eine spätere Region  $n+\dots$  fällt. Die Total Reading Time gibt die Summe aller Fixationen auf einer Region  $n$  wieder. Die Regression Path Duration berechnet die Summe aller Fixationen auf einer Region  $n$  sowie auf allen anschließend fixierten vorhergehenden Regionen  $n-\dots$ , bevor ein Blicksprung über die Region  $n$  in eine spätere Region  $n+\dots$  fällt. Die Second Pass Reading Times schließlich berechnen die Differenz der Total Reading Times und der First Pass Reading Times – wobei hierfür auch andere Berechnungsmethoden zum Einsatz kommen. Den verschiedenen Maßen werden Beteiligungen an verschiedenen Verarbeitungsprozessen beigemessen. Dabei ist aber mitunter unklar, welche Prozesse durch welche Maße reflektiert werden. First Fixations und First Pass Reading Times werden zumeist mit frühen Verarbeitungsprozessen assoziiert, Second Pass Reading Times hingegen eher mit späteren Integrationsprozessen.

Ähnlich wie im zweiten Experiment von Tabor et al. (2004) wurden im Experiment von Müller et al. (2009) Konstruktionen zugrundegelegt, in denen Argumente in Objektposition stehen, die in der globalen Interpretation unproblematisch sind, innerhalb der isolierten lokalen Interpretation aber semantisch unplausibel sein müssen. Somit werden neben den Kontrollbedingungen ohne LSK auch zwei LSK-Bedingungen variiert: Semantisch plausible LSK (*Der Chef überzeugt den Kollegen aus der PR-Abteilung.*, Bsp. 12a) und semantisch unplausible LSK (*Der Chef überzeugt den Schreibtisch aus der PR-Abteilung.*, Bsp. 12). Sollten LSK unmittelbar semantisch verarbeitet werden, so sollten sich im Falle der lokalen semantischen Verletzung entsprechende Effekte mit höheren Lesezeiten finden. Die Effektrichtung der LSK betreffend zeigen sich während der Verarbeitung der Präpositionalphrase (*aus der PR-Abteilung*) die niedrigsten First-Pass-Lesezeiten für Bedingung 1, d.h. während der LSK. Zugleich finden sich für diese Bedingung die höchsten First-Pass- und Regression-Path-Lesezeiten auf dem Relativsatz-Verb *zuteilt*. Müller et al. (2009) interpretieren dies als Hinweis darauf, dass es während der LSK zu einer temporären Erleichterung der Verarbeitung sowie nach dem Ende der LSK am Punkt des Zusammenbrechens der global inkonsistenten lokalen Interpretation zu einer Erschwerung der Verarbeitung kommt – entsprechend der Vorhersage von Ansätzen unbeschränkter Parallelverarbeitung. Die erwartete Erhöhung der Lesezeiten auf der lokal semantisch unplausiblen Nominalphrase *den Schreibtisch* in Bedingung 3 wurde nicht gefunden. Allerdings finden sich hier im Unterschied zu Bedingung 1 auch keine Effekte der LSK. Die Verarbeitung von LSK scheint also nicht alleine von der syntaktischen Kohärenz der lokalen Sequenz abhängig zu sein, sondern auch von ihrer semantischen Kohärenz. Dies weist auf eine unmittelbare semantische Verarbeitung hin, die bereits im lokalen Kontext beginnt.

- (12) a. Wilfried, dem der Chef überzeugt den Kollegen aus der PR-Abteilung zuteilt, geht früh nach Hause.  
 b. Wilfried, dem der Chef entschlossen den Kollegen aus der PR-Abteilung zuteilt, geht früh nach Hause.

- c. Wilfried, dem der Chef überzeugt den Schreibtisch aus der PR-Abteilung zuteilt, geht früh nach Hause.
- d. Wilfried, dem der Chef entschlossen den Schreibtisch aus der PR-Abteilung zuteilt, geht früh nach Hause.

**Einflüsse interindividueller Differenzen** Ein weiterer Aspekt, der bei all diesen Untersuchungen immer wieder eine Rolle spielt, ist die Frage des Einflusses interindividueller Differenzen. In dieser Hinsicht stehen die Untersuchungen in Beziehung zu den Befunden und Annahmen von MacDonald und Christiansen (2002) und Wells et al. (2009) sowie zur Replikation durch Konieczny, Ruh und Müller (2009) und ihrer Interpretation (Konieczny & Müller, 2010). In all diesen Studien sind die Annahmen aus Ergebnissen von SRN-Simulationen begründet. In diesem Zusammenhang wird ein starker Einfluss der Erfahrung auf die Leistungsfähigkeit bei der Sprachverarbeitung abgeleitet (vgl. auch Kap. 3.2.2 und 3.2.3). Dabei stellt sich für LSK die Frage, welchen Einfluss die verbale Leistungsfähigkeit auf diesen Gegenstand hat – sei es hinsichtlich sprachlicher Erfahrung oder größerer Verarbeitungsressourcen – und wie sich diese mit den Vorhersagen verschiedener Sprachverarbeitungsansätze und Leistungsmaße (wie etwa der Lesespanne) decken.

Konieczny (2005) erläutert entsprechende Gesichtspunkte in erfahrungsbasierten Modellen. Hier spielen die Frequenzen und Regularitäten der verschiedenen Strukturen und sprachlichen Einheiten, die auf einer bestimmten Satzposition zum Tragen kommen, eine bedeutende Rolle. Dazu zählen unter anderem: die globale Struktur, die Teilsatzstruktur, verschiedene auftretende LSK und die Frequenz ihrer Struktur, wenn sie als global konsistente Struktur isoliert auftreten würden, die Frequenz ambiger lexikalischer Elemente und die Frequenzen der syntaktischen Verwendung ihrer verschiedenen Interpretationsmöglichkeiten. Diese Frequenzen müssen gegeneinander abgewogen werden. Eine höhere Sprach- und Leseerfahrung mit komplexem Sprachmaterial könnte dazu führen, dass das Fenster der berücksichtigten Struktur und der strukturellen Zusammenhänge erweitert wird, d.h. es wird bei der Verarbeitung ein größerer lokaler Kontext berücksichtigt. Dadurch können Abhängigkeiten und strukturelle Bedingungen über größere Distanzen hinweg besser verfolgt werden und gewinnen damit auch bei den Abwägungen der Aufmerksamkeitsverteilung an Bedeutung, während kleinere LSK besser ausgeschlossen werden können. Personen mit besseren sprachlichen Fähigkeiten sollten also eine bessere Performanz für komplexere und unregelmäßige Strukturen zeigen und weniger anfällig für LSK sein. In diesem Zusammenhang wurde auch Inhibition als entscheidender kognitiver Faktor diskutiert: Erfahrenere Leser und Hörer können möglicherweise während des Sprachverstehens kleinere lokale Kohärenzen besser unterdrücken (vgl. auch Hachmann, 2008; Hachmann, Konieczny & Müller, 2009).

Um weiteren Aufschluss über die Effekte zu erhalten, wurden in die Studien mitunter Erhebungen zur Spracherfahrung und Experimente zur Bemessung des sprachlichen Arbeitsgedächtnis aufgenommen, um Haupteffekte der Gruppeneinteilung in verschiedene Lese- und Sprachkompetenzgruppen und Interaktionen mit den Experimentalfaktoren berücksichtigen zu können. So zeigt Konieczny (2005) Unterschiede zwischen Gruppen verschiedener Erfahrung bei der Verarbeitung von LSK. Allerdings lassen die Ergebnisse keine definitive Aussage zu, welcher Erklärungsansatz favorisiert werden kann. Auch aus den anderen Studien sind bislang keine distinktiven Ergebnisse zu dieser Frage bekannt.

### 5.2.3. Probleme und offene Fragen

**Kontrolle prosodischer Effekte** Welche Rolle spielt die Markierung satzinterner Satzteilgrenzen – prosodische Marker oder Interpunktionszeichen – für die Ausprägung und die Stärke von LSK-Effekten? Kommata markieren im Deutschen Übergänge in neue Teilsätze und begrenzen abgeschlossene Nebensätze im nicht abgeschlossenen Matrixsatz (z.B. zentraleingebettete Relativsätze). Prosodische Marker wie Pausen zwischen Teilsätzen, tonale Hebungen am Satzende von Nebensätzen im nicht abgeschlossenen Matrixsatz oder tonale Senkungen am abschließenden Satzende klären strukturelle Einheiten und Zusammenhänge in der gesprochenen Sprache. Auf einen entscheidenden und unmittelbaren Einfluss auf die strukturelle Interpretation der Prosodie beim Verstehen gesprochener Sprache weisen Befunde aus verschiedenen experimentellen Paradigmen hin (vgl. z.B. DeDe, 2010; Steinhauer, Alter & Friederici, 1999; Price, Ostendorf, Shattuck-Hufnagel & Fong, 1991).<sup>8</sup>

Die bisher angeführten Studien unterscheiden sich darin, ob Prosodie oder Interpunktion die Verarbeitung der lokalen Subsequenz beeinflussen könnten oder nicht: In der Lesezeitmessung von Tabor et al. (2004) mit Self-paced Reading überschreitet die LSK die Satzgrenze zwischen Matrixsatz und reduziertem Relativsatz, allerdings wird dieser hypotaktische Übergang im Englischen nicht wie im Deutschen durch Kommata markiert. Der Effekt wird also nicht durch künstliche Maßnahmen forciert, wenngleich die Frage besteht, ob diese Ergebnisse ohne Weiteres auf gesprochene Sprache übertragen werden können, denn hier würde eine prosodische Markierung vorliegen. Genau dieses Problem der reichen prosodischen Struktur gesprochener Sprache muss in den Studien im Visual-World-Paradigma kontrolliert werden. In der Studie von Konieczny, Müller, Hachmann et al. (2009) wird durch die LSK zwar keine Satzgrenze überschritten, da sie sich komplett innerhalb des Komplementsatzes befindet. Dennoch kann nicht ausgeschlossen werden, dass das morphologisch ambige Partizip prosodisch als Partizip markiert wird. Dasselbe trifft für die Studie mit Dativrelativsatz zu (Konieczny, Müller, Baumann et al., 2009). Noch stärker tritt die prosodische Markierung im Stimulusmaterial zu Tage, in dem die LSK die Satzgrenze zwischen Relativsatz und fortgesetztem Matrixsatz überschreitet (ebenfalls in Konieczny, Müller, Baumann et al., 2009). Hier findet sich auf dem satzfinalen Relativsatz-Verb eine steigende Kontur der Intonation sowie eine anschließende Pause. Diese Markierung dürfte den Effekt der LSK unterbinden, denn sie durchbricht ganz klar die prosodischen Merkmale eines geschlossen kohärenten Teilsatzes. Zudem muss durch die lokale prosodische Markierung gar nicht die globale Struktur des Satzes berücksichtigt werden – die lokal durch die Intonation verfügbare Information genügt, um die Interpretation der LSK auszuschließen.

Um diesen Einfluss zu kontrollieren, verwenden Konieczny und Kollegen prosodisch manipuliertes Satzmaterial: Dazu wurden die natürlich eingesprochenen Sätze mithilfe von Audiotbearbeitungsprogrammen verändert, indem zusätzliche, prosodisch möglichst neutral gesprochene Sequenzen der relevanten Passagen eingeschnitten wurden (vgl. z.B.

<sup>8</sup>Allerdings sind die Ergebnisse von DeDe (2010) mit Vorbehalten zu betrachten, da als Paradigma der Studie Self-Paced Listening (mit auditiver Repräsentation des Satzes Wort für Wort durch Tastendruck) verwendet wurde. Dies entspricht keinem natürlichen Sprachverstehen und verzerrt zudem viele prosodische Merkmale wie etwa Pausen.

Konieczny & Müller, 2007; Konieczny, Müller, Hachmann et al., 2009; Konieczny, Müller, Baumann et al., 2009; Weldle, Konieczny, Wolfer et al., 2010). Dabei wurden die Kontrollbedingungen mit demselben Material mit eingeschnittenen Sequenzen erstellt. Auf diese Weise wurden prosodische Hinweise eliminiert und zudem eine größere Homogenität der unterschiedlichen Satzbedingungen erreicht. Durch die identische Manipulation sollte sichergestellt werden, dass die gefundenen Effekte durch die LSK zustande kommen und nicht durch prosodische Unterschiede der Bedingungen. Durch diese Methodik zur Kontrolle des Einflusses der Prosodie wird erreicht, dass die Verarbeitung der Sätze in erster Linie aufgrund lexikalischer und syntaktischer Informationen erfolgt.<sup>9</sup>

Unter diesen Bedingungen wurden LSK-Effekte zum einen für gesprochene Sprache und zum anderen bei Überschreitung von Satzgrenzen gefunden. Doch inwiefern entspricht diese Performanz für reduziertes auditives Sprachmaterial natürlicher Satzverarbeitung? Zumindest muss berücksichtigt werden, dass das Fehlen entsprechender Marker das natürliche Zusammenwirken verschiedener Constraints und Mechanismen einschränkt und ihre Gewichtung im Analyseprozess in ein künstliches Ungleichgewicht versetzt, und damit alternative Präferenzen oder Strategien begünstigt werden. Um entsprechende Unterschiede zu erfassen, kontrastieren Konieczny, Müller, Baumann et al. (2009) auch den Effekt der Prosodie: So zeigte sich, dass der Effekt der LSK nicht mehr reliabel ist, wenn bei identischem Design der Studie natürlich gesprochene Satzmaterialien verwendet werden. Dies ist ein weiterer Hinweis auf den unmittelbaren Einfluss der Prosodie auf die syntaktische Verarbeitung. Wir müssen also davon ausgehen, dass die Auswirkungen von LSK im alltäglichen Sprachgebrauch vor allem durch lokale prosodische Hinweise kontrolliert wird – Prosodie unterstützt die korrekte Unterteilung und Zuweisung syntaktischer Strukturen. Zugleich zeigt diese Rolle der Prosodie, in welchem Maße die korrekte syntaktische Verarbeitung abhängig von ihr ist, wie fragil die isolierte syntaktische Verarbeitung in Wirklichkeit ist und wie lokal sie erfolgt.

**Visuelle Induzierung von Effekten** Inwiefern werden die Effekte im Visual-World-Paradigma, die auf die Verarbeitung von LSK hinweisen, durch die visuelle Abbildung der zur LSK korrespondierenden Szene induziert? Diese Frage betrifft die Unterscheidung von Ursache und Wirkung im verwendeten Paradigma bei der Blickbewegungsmessung. So wäre es möglich, dass die Abbildung der verschiedenen Interpretationsmöglichkeiten (in den Studien von Konieczny & Müller, 2007; Konieczny, Müller, Hachmann et al., 2009; Konieczny & Müller, 2010; Weldle, Konieczny, Müller et al., 2010a) die Inbetrachtung dieser Interpretationsmöglichkeiten beeinflusst. Beim Versuch, den gehörten Satz auf die visuell angebotenen Interpretationsmöglichkeiten abzubilden bzw. visuelle und linguistische Informationen zu integrieren, könnten also die visuellen Informationen Interpretationen

---

<sup>9</sup>In einer unveröffentlichten Rating-Studie haben wir die auditiven Materialien mit und ohne prosodische Manipulation verglichen, um festzustellen, ob die Manipulation die Natürlichkeit oder die Verständlichkeit für die Satzmaterialien beeinflusst (Details zur Studie finden sich in Kap. 7.3.6.1). Für die Version mit Dativ-Relativsätzen zeigte sich kein Unterschied im Rating für geschnittene und natürliche Sätze. Für die Version mit Komplementsätzen hingegen wurde ein Unterschied gefunden: Die manipulierten Sätze wurden schlechter bewertet. Da die Manipulation aber für Ziel- und Kontrollsätze gleichermaßen durchgeführt wurde, findet sich für die Experimentalbedingungen die gleiche Ausgangslage – gefundene Effekte beruhen also auf experimentell kontrollierten Kontrasten und nicht auf Unterschieden der prosodischen Natürlichkeit.

nahelegen, die ohne die visuelle Information überhaupt nicht oder nur marginal in Betracht gezogen werden würden – zum Beispiel aufgrund der identischen Agenten. Die visuelle Information würde damit Mechanismen und Constraints der linguistischen Verarbeitung beeinflussen.

Auf solche Effekte des frühzeitigen Einflusses des visuellen Kontextes auf die Sprachverarbeitung weisen unter anderem Tanenhaus et al. (1995) hin: In ihrer Studie finden sich Hinweise darauf, dass bei temporären Anbindungsambiguitäten (Präpositionalphrase vs. reduzierter Relativsatz, vgl. Bsp. 13 mit Markierung der ambigen Position durch Doppel-Asterisk) frühzeitig der visuelle Kontext zur Auflösung der Ambiguität bzw. zur Bildung von Anbindungspräferenzen herangezogen wird (vgl. auch Eberhard et al., 1995). Im Beispiel wird die Relativsatz-Lesart (mit reduziertem Relativsatz, Bsp.13b) gegenüber der direktionalen Lesart (13c) präferiert, wenn die visuelle Lesart eine Objekt-Diskriminierung durch einen Relativsatz nahelegt (etwa durch Abbildung zweier Äpfel, von denen einer auf einem Handtuch und ein anderer in einer Schachtel liegt).

- (13) a. Put the apple \*\*on the towel.  
 b. Put the apple that is on the towel in the box.  
 c. Put the apple \*\*on the towel in the box.

Entsprechend könnten die LSK-Effekte nur deshalb erscheinen, weil die Möglichkeit der Korrespondenz zwischen Satz und Abbildung der LSK-Situation durch den visuellen Kontext nahegelegt wird. Diese Möglichkeit kann innerhalb der Studien im Visual-World-Paradigma mit szenischen Abbildungen nicht endgültig entkräftet werden. Was dennoch bleibt, ist der Effekt der kontrastierten experimentellen Bedingungen, die selbst im Falle visueller Induzierung auf eine temporäre, stärkere Inbetrachtung der LSK-Interpretation in der LSK-Bedingung hinweist. Abgesehen davon unterstützen aber die kombinierten Resultate verschiedener Studien die Annahme, dass LSK interpretiert werden. So finden sich ebenfalls Effekte in Studien, die andere Experimental-Paradigmen verwenden, zum Beispiel Self-Paced Reading (Tabor et al., 2004) oder Blickbewegungsmessungen während des Lesens (Müller et al., 2009). In beiden ist die Beeinflussung der Sprachverarbeitung durch einen visuellen Kontext nicht möglich, da nur linguistische Informationen gegeben sind. Allerdings kann anhand dieser Studien wiederum nicht endgültig klargestellt werden, ob LSK tatsächlich Gegenstand der Interpretation werden oder nur eine Quelle von Interferenzen gegen die globale Interpretation darstellen. Weitere Hinweise auf die Interpretations-Hypothese finden sich innerhalb des Visual-World-Paradigmas: In einer Reihe von Studien zur Verarbeitung anaphorischer Referenzen innerhalb von LSK wurden entsprechende Effekte gefunden (diese Studien sind Gegenstand des Kapitels 7). Diese können nicht auf eine visuelle Induzierung zurückgeführt werden, da in ihnen für die visuellen Stimuli ausschließlich Abbildungen kontextfreier Einzelreferenten verwendet werden, die keine Interpretation nahelegen (Weldle, Konieczny, Wolfer et al., 2010).

**Offene Details der Effekte** Weitere offene Fragen betreffen die Details der Effekte und die Vorhersage ihrer Auswirkung auf neue Strukturen. Es ist nach wie vor nicht endgültig geklärt, welche Effekte für LSK in den verschiedenen Experimental-Paradigmen aus welchen Gründen erwartet werden sollen. Dies hängt natürlich stark von der Frage ab,

welches Sprachverarbeitungsmodell mit welchen Eigenschaften zugrundegelegt wird – ob modulare oder interaktive Verarbeitung verschiedener Informationsaspekte, serielle oder parallele Verarbeitung verschiedener Parsingalternativen und freie oder beschränkte Ressourcen der Parallelverarbeitung angenommen werden.

Während die erwarteten Effekte im Visual-World-Paradigma vergleichsweise klar formuliert werden können und bereits recht gut untersucht und belegt sind, ist dies bei den Lesezeitmessungen in den Paradigmen Self-Paced Reading und Blickbewegungsaufzeichnung während des Lesens sowie in Offline-Maßen (etwa Rating-Studien oder Verständnisfragen) eher unklar und mitunter können sogar entgegengesetzte Auswirkungen erwartet werden. Erstens stellt sich dabei die Frage, welche Ausprägungen die Effekte haben – Erleichterung oder Erschwerung der Verarbeitung bzw. Vergrößerung oder Verkleinerung der Lesezeit. Zweitens ist zu klären, auf welchen Positionen welche Effektausprägung erwartet wird – etwa auf der Position des Auftretens der LSK, im Verlauf und am Ende der LSK, auf der Position des Zusammenbrechens bzw. Scheiterns der lokalen Struktur sowie auf den darauffolgenden Positionen, während derer eventuell die globale Interpretation wieder etabliert wird. Drittens ist unklar, welche Auswirkung LSK auf die langfristige Repräsentation des Satzinhalts haben – auf die Beantwortbarkeit von Fragen, die sich auf Propositionen beziehen, die hinsichtlich ihres semantischen Inhalts durch die LSK betroffen sind sowie auf die Beurteilung der Grammatikalität oder Plausibilität des Gesamtsatzes.

Die bisherigen Befunde für die Lesezeiten sind mitunter widersprüchlich und enthalten Resultate entgegengesetzter Ausprägung auf identischen Positionen: So finden Tabor et al. (2004) erhöhte Lesezeiten während der LSK. Am Ende der LSK bzw. auf der nachfolgenden Position hingegen zeigt sich kein Effekt (Studie 1). Zudem finden sie sowohl für die semantisch plausible (belebtes Subjekt der LSK) als auch für die semantisch unplausible LSK (unbelebtes Subjekt der LSK) Effekte, wobei der Effekt für die plausible Variante stärker ausfällt. Beinahe das genaue Gegenteil findet sich in den Ergebnissen von Müller et al. (2009): Hier zeigen sich geringere Lesezeiten während der LSK. Zudem findet sich ein deutlicher Effekt nach Ende der LSK: Auf der Position, auf der die lokale Struktur verworfen werden muss, finden sich deutlich erhöhte Lesezeiten. Und auch die Ausprägung der semantischen Plausibilität innerhalb der LSK verhält sich anders, denn in dieser Studie zeigt sich hierfür kein Effekt mehr. Ein möglicher, wenngleich nicht besonders überzeugender Grund für diese widersprüchlichen Ergebnisse könnte sein, dass die Studien mit unterschiedlichen syntaktischen Strukturen und zudem in unterschiedlichen Sprachen durchgeführt wurden. Dies schränkt die Vergleichbarkeit ein und erschwert durch die einzelsprachspezifischen Strukturen, innerhalb derer sich LSK befinden können, auch in Zukunft die sprachübergreifende Untersuchung. Dennoch kann dies kaum eine Erklärung für diese starke Abweichung der Resultate sein. Eine schlüssige Erklärung zu finden ist allerdings schwierig und wird nur über weitere, spezifischere Untersuchungen zu klären sein.

Auch für die Verarbeitungstiefe und -beständigkeit für LSK (d.h. welche linguistischen Ebenen werden durch das Phänomen berührt) werden in den neueren Studien zwar erste Antworten gefunden, dennoch bleiben viele Fragen offen und die Befunde sind noch recht unspezifisch. So wird die Verarbeitung von LSK auf der semantisch-interpretatorischen Ebene attestiert, offen bleibt aber die Tiefe der semantischen Verarbeitung. Wird etwa aufgrund der LSK die Antizipation der kommenden Argumentstruktur, der erwarteten

Argumente und der präferierten semantischen Merkmale der Argumente beeinflusst? Die Frage der Verarbeitungstiefe betrifft auch die Gewichtung der Verarbeitung von LSK und damit auch die Unterscheidung zwischen Ansätzen paralleler und serieller Verarbeitung. Hier beispielhaft bezogen auf die Befunde aus den Visual-World-Daten: Erfolgt die Verarbeitung im Sinne serieller Verarbeitung absolut, so dass sich manche VTn in manchen Fällen komplett für die Interpretation der LSK und gegen die globale Interpretation entscheiden, und dementsprechend jeweils entweder die zur globalen oder zur lokalen Interpretation korrespondierende Abbildung fixieren? Oder erfolgt sie im Sinne paralleler Verarbeitung relativ, d.h. sowohl die globale als auch die lokale Interpretation werden kurzfristig gleichzeitig verfolgt, was zu einer erhöhten Wahrscheinlichkeit des Springens der visuellen Aufmerksamkeit zwischen den beiden korrespondierenden Abbildungen führt? Beide Möglichkeiten dürften in der Analyse der Fixationswahrscheinlichkeiten oberflächlich in etwa denselben Ausprägungen der Effekte resultieren. Ebenso unklar bleibt bislang die Persistenz der Interpretation aus der Verarbeitung der LSK. Bleibt die Interpretation in einer Stärke und Beständigkeit im mentalen Situationsmodell aufrechterhalten, so dass sie auch bei Verständnisfragen oder bei der Abfrage von (propositionalen) Fakten nachhaltig die Performanz beeinflusst? In den Studien von Tabor et al. (2004) wurden bereits geschlossene Verständnisfragen gestellt – leider wurden aber weder deren Aufbau und Fragegegenstand noch die Resultate für die Fragen dargelegt. Ansonsten wurde die Persistenz der LSK-Interpretation bislang weitgehend ausgespart. Folglich werden hier noch einige Punkte zu klären sein.

### 5.2.4. Theoretische Erklärungsansätze

Das Phänomen LSK scheint den bisherigen Befunden nach vor allem temporär Einfluss zu nehmen. Zudem weisen die Fixationsanteile aus den Visual-World-Studien darauf hin, dass die Fixationen korrespondierend zur globalen Interpretation nach wie vor den Hauptanteil der Aufmerksamkeit stellen. Weshalb also sollte die Beschäftigung damit von solchem Interesse sein, wenn es doch auf den ersten Blick nur eine unter vielen anderen Interferenz-Quellen ist? In der Tat stellen LSK ein wesentlich bedeutenderes Phänomen für die Identifizierung der Prozesse natürlicher Sprachverarbeitung dar, und das selbst dann, wenn wir es im alltäglichen Gebrauch nicht einmal bemerken würden oder es möglicherweise sogar gänzlich durch prosodische Informationen komplett abgefangen wird. Denn die Tatsache, dass lokale Subparses überhaupt einen Einfluss nehmen, obwohl sie im globalen Kontext keine Rolle spielen dürften und sofort als ungrammatische Alternative verworfen werden müssten, stellt ein Problem für den größten Teil der gängigen Theorien und Modelle der Satzverarbeitung dar. Der Grund dafür ist die Einschränkung der Annahme einer global konsistenten Interpretation von Sätzen, in der bei der inkrementellen Integration neuer Informationen stets die gesamte, bis dahin aufgebaute Informationsstruktur des Satzes berücksichtigt wird.

Die Annahme globaler Konsistenz ist für viele Ansätze axiomatisch, wird aber durch den Befund von LSK in Frage gestellt. In der Folge können wir Grundannahmen und Eigenschaften verschiedener Ansätze zur Satzverarbeitung hinsichtlich ihrer Erklärung

des Phänomens und hinsichtlich ihrer Vorhersage der Verarbeitungsperformanz für dieses Phänomen vergleichen und kritisch überprüfen.<sup>10</sup>

### 5.2.4.1. Dynamische und unvollständige Verarbeitung

**Dynamische Modelle** Konieczny und Kollegen sehen die Befunde als Argument für Modelle, die unbeschränkte Parallelverarbeitung gewährleisten und in denen die Annahme globaler Konsistenz nicht zwingend gegeben ist bzw. erfüllt wird (Konieczny, 2005; Konieczny, Müller, Hachmann et al., 2009; Konieczny & Müller, 2010). Als geeignetster Ansatz hierfür werden dynamische Modelle vorgeschlagen. Diese Annahme beruht auf Modellierungen mit konnektionistischen Modellen, speziell SRNs, die entsprechende dynamische Eigenschaften aufweisen. Sie decken sich damit gut mit neueren Befunden der psycholinguistischen Forschung, und erfassen neben LSK die Gegebenheit von Lokalitäts- und Antilokalitätseffekten innerhalb desselben Modells. Speziell mit SRNs wurden LSK-Effekte simuliert (vgl. Kapitel 3.2.3) und Erklärungsansätze zum Phänomen abgeleitet (vgl. Konieczny & Ruh, 2003; Konieczny, Ruh & Müller, 2009): False Alarms (d.h. Aktivierungen falscher Wortklassen im Sinne fehlerhafter Antizipation der syntaktischen Struktur) entstehen durch Interferenzen konkurrierender, lokal konsistenter Subsequenzen, die durch die beschränkte Kapazität des Systems bzw. durch die Überlagerung von Informationen und die dadurch begünstigte Dominanz von Abhängigkeiten im lokalen Aufmerksamkeitsfenster verursacht werden. Andere Interpretationen im Ansatz dynamischer Modelle heben den Einfluss kompetitiver Attraktoren auf die Erreichung eines stabilen Zustands der Parsingstruktur hervor, so etwa im VSG-Modell (Visitation Set Gravitation, Tabor et al., 1997; Tabor & Tanenhaus, 1999) oder im SOPARSE-Modell (Self-Organized Parsing, Tabor & Hutchins, 2004). In weiteren Modellen basieren die Effekte auf der interaktiven Aktivierung lexikalischer Frames und dem Wettbewerb durch Inhibition, etwa im Unification Space-Parser (Vosse & Kempen, 2000) sowie in dessen erweiterter lokalistisch-konnektionistischer Implementierung im SINUS-Modell (Simple/Semi Interactive Activation and Competition Neural Unification Space Network, Vosse & Kempen, 2009).

Ein großer Vorteil dynamischer Modelle für die Erklärung von LSK-Effekten liegt darin, dass das Phänomen als Nebenprodukt inhärenter Eigenschaften der Modelle bei der normalen Verarbeitung entsteht (Konieczny, Müller, Hachmann et al., 2009). Damit ergibt sich ein wesentlicher Unterschied zu klassisch-symbolischen, rationalen Modellen der Satzverarbeitung. In diesen wäre die Verarbeitung von Strukturen, die unter der Annahme globaler Konsistenz bereits vor ihrer Verarbeitung als ungrammatisch ausgeschlossen werden können, ein irrationales Verhalten (Konieczny & Müller, 2010, S. 48):

Im Rahmen eines symbolisch/algorithmischen Ansatzes erscheint es irrational, dass wir Ressourcen verschwenden für Analysen, die von vornherein ausgeschlossen werden können. Wir interpretieren den Effekt lokaler Kohärenzen als Hinweis auf einen nicht-rationalen Nebeneffekt der zugrundeliegenden Architektur des Spracherwerbs- und Sprachverarbeitungssystems.

---

<sup>10</sup>Eine detailliertere Unterscheidung der verschiedenen Erklärungsansätze hinsichtlich ihrer Aussagen und Vorhersagen für Aspekte der Verarbeitung von LSK ist einer der zentralen Gegenstände der aktuell an der Universität Freiburg entstehenden Dissertation von Daniel Müller mit dem Thema „Modeling frequency effects in local syntactic coherence processing“.

Um solche Phänomene abzudecken, müssten dieser Einschätzung folgend in rationalen Modellen zusätzliche Eigenschaften oder Mechanismen des Modells postuliert werden, für die es keine offensichtliche Rechtfertigung gibt, da sie einerseits die Verarbeitungsressourcen betreffend kostenintensiv sind, umgekehrt aber keinen sinnvollen Nutzen in einem anderen Bereich eröffnen und damit schlicht unökonomisch sind. Umgekehrt beruhen Interferenz-Effekte des lokalen Kontexts in unbeschränkt parallel verarbeitenden Ansätzen, wie zum Beispiel konnektionistischen Netzwerken, durchaus auf ökonomischen Eigenschaften, denn „[...] die Satzverarbeitung [profitiert] in der überwältigenden Mehrheit der Fälle dramatisch durch die Möglichkeit, die nächsten Wörter auf der Basis von lokalen Wahrscheinlichkeiten vorherzusagen.“ (Konieczny & Müller, 2010, S. 49). Dies ist ein vorteilhafter Einsatz für die vergleichsweise eher seltene Komplexitätszunahme durch LSK.

**Beschränktes und partielles Parsing** Im Zusammenhang mit Erwägungen zur Ökonomie wurde die Interpretation dieses (begrenzt rationalen) dynamischen Ansatzes mit dem Ansatz des Good-Enough-Parsing (Ferreira et al., 2002; Ferreira & Patson, 2007) in Verbindung gebracht. In diesem wird die Annahme kompositionell-algorithmischer Ansätze des Sprachverstehens mit vollständigen, detaillierten und akkuraten Repräsentationen des linguistischen Inputs anhand widersprüchlicher empirischer Beobachtungen problematisiert. In ihrem Gegenentwurf schlagen Ferreira et al. (2002) vor, dass die in der Verarbeitung vom Input abgeleitete Bedeutung nicht der wahren Bedeutung entspricht und auch selbst nach Reanalyse noch inkorrekt sein kann. Anstatt vollständig und global konsistent zu sein, erfolgt die Verarbeitung nur partiell und die abgeleitete Repräsentation des sprachlichen Inputs ist unvollständig – Sprachverarbeitungsprozesse sind also nicht auf Perfektion ausgelegt, sondern eben nur ‚gut genug‘, um das Verstehen in den allermeisten Situationen zu gewährleisten und dabei möglichst schnell und robust zu erfolgen. In diesem Zusammenhang wurde vorgeschlagen, dass SRNs als Implementierung des Good-Enough-Parsing verstanden werden können (Konieczny, Müller, Hachmann et al., 2009).

Alternativ zu SRNs und zu anderen Ansätzen zu LSK (siehe unten) wurde eine Implementierung des Good-Enough-Ansatzes mit einem Bottom-up-Parser auf Basis einer Combinatory Categorical Grammar vorgeschlagen (Morgan, Keller & Steedman, 2010). Durch das Primat der Bottom-up-Verarbeitung wird die Verwendung ausschließlich lokaler Informationen erzwungen. Dabei soll der Grammatikformalismus sicherstellen, dass die Inkrementalität des Parses gewährleistet ist. Die Schwierigkeit der Verarbeitung eines Satzes wird dabei als Verhältnis der Wahrscheinlichkeit bzw. Plausibilität der in ihm enthaltenen LSK zur Plausibilität der korrekten syntaktischen Interpretation. Auf dieser Grundlage können mit dem Modell Werte für Grammatikalitätsratings auf Satzebene vorhergesagt werden, wie für die Ratings aus der Studie von Tabor et al. (2004) demonstriert wurde. Eine Wort-für-Wort-Zuweisung der Komplexität ist in diesem Ansatz allerdings nicht möglich, womit seine Umsetzbarkeit für die psycholinguistische Forschung und seine Aussagekraft eher beschränkt bleiben.

Ein möglicher alternativer Erklärungsansatz für LSK-Effekte, der ebenfalls eine gewisse Nähe zur Idee des Good-Enough-Parsing aufweist, ist Shallow Parsing, auch Partial Parsing oder Chunking genannt (eine Einführung zum Thema im Special Issue des *Journal*

of *Machine Learning Research* geben Hammerton, Osborne, Armstrong & Daelemans, 2002). Dieser Ansatz hat vor allem Relevanz für Ansätze der maschinellen Verarbeitung natürlicher Sprache, wurde aber ursprünglich hinsichtlich seiner Relevanz zur Erklärung psycholinguistischer Befunde vorgeschlagen (vgl. Abney, 1989). Shallow Parsing kann als Vorstufe zum vollständigen Parsing eines Satzes verstanden werden, bei der noch keine Erfassung der globalen Tiefenstruktur des Satzes und seiner semantischen Interpretation erfolgt: Es geht dabei um das Identifizieren und Unterteilen von einfachen Phrasen im Text im Sinne einer beschränkten Analyse grundlegender syntaktischer Eigenschaften von Subsequenzen. In der maschinellen Anwendung bedeutet dies, mithilfe regelbasierter oder statistischer Verfahren partielle Analysen eines Satzes zur Erfassung lokal zusammengehöriger Subsequenzen zu leisten. Diese Leistung kann im Vergleich zum Tiefenparsing sparsam, effizient und robust erfüllt werden. Denkbar ist, dass die Strukturen, in denen wir LSK finden bzw. bisher untersucht haben, so komplex sind, dass mitunter das Tiefenparsing misslingt und statt einer Reanalyse auf die Ergebnisse des Shallow Parsing-Prozesses zurückgegriffen wird – sofern ein solcher Prozess als Vorstufe zum eigentlichen Parsing tatsächlich existiert. Die Effekte der Interferenz oder sogar Interpretation von LSK entstehen aus dem Versuch, den Satz auf der Grundlage dieser unvollständigen syntaktischen Informationen zu interpretieren.

### 5.2.4.2. Symbolische und rationale Ansätze

Den dynamisch-konnektionistischen Modellen stehen theoretische Ansätze und Modelle gegenüber, die eine symbolisch-algorithmische bzw. rationale Perspektive beibehalten. Diese müssen eine auf rationalen Prinzipien beruhende Erklärung dafür finden, warum in bestimmten Kontexten nicht alle verfügbaren Informationen des globalen Kontexts genutzt werden, und weshalb der Parsingprozess von Strukturen beeinflusst wird, die einen nicht-rationalen Zusatzaufwand erfordern, obwohl sie auf den kritischen Positionen bereits als Analyse ausgeschlossen sein sollten. Zur Erklärung werden verschiedene Effekte herangezogen: Effekte des Einflusses lexikalischer Elemente (statistische Eigenschaften der zu integrierenden Elemente, Konkurrenz ähnlicher Elemente), Effekte der Differenz struktureller Wahrscheinlichkeiten des Inputs im Verhältnis zum Gesamtsatz (Abgleich zwischen Bottom-up- und Top-down-Informationen) und nicht zuletzt Effekte durch Ressourcen-Beschränkungen (v.a. des Arbeitsgedächtnisses). Vielfach wird dabei anhand der Auswirkungen der Ambiguität argumentiert, die die lokale Interpretationsmöglichkeit der LSK verursacht. Allerdings ist mitunter unklar, wie sich in diesen Ansätzen LSK-Effekte erklären, die nicht auf lexikalischen Ambiguitäten beruhen (so z.B. in Konieczny, Müller, Baumann et al., 2009; Weldle, Konieczny, Wolfer et al., 2010).

**Statistische lexikalische Erwartungen** Gibson (2006) schlägt in seinem Ansatz kontext-unabhängiger Kategoriefrequenzen vor, dass Verarbeitungspräferenzen aus der Interaktion zwischen lokalen statistischen Bottom-up-Informationen zur relativen Häufigkeit möglicher syntaktischer Kategorien lexikalischer Einheiten (kontext-unabhängig) und global-syntaktischen statistischen Top-down-Erwartungen (kontext-abhängig) resultieren. Auf dieser Grundlage werden Ambiguitäten der syntaktischen Kategorie lexikalischer Einheiten aufgelöst. Mit diesem Ansatz lassen sich mit Hilfe von Verteilungstatistiken aus

linguistischen Korpora formal Erwartungswerte für Lesezeiten berechnen und Verarbeitungspräferenzen abschätzen.<sup>11</sup> Gibson klassifiziert diesen Ansatz als selbst-konsistentes Parsing und stellt ihn den Ansätzen lokaler Kohärenz gegenüber. Er diskutiert dies kontrastierend vor allem auf die Studie von Tabor et al. (2004) bezogen, und zeigt, wie die Effekte auf Grundlage statistischer Verteilungen einzelner lexikalischer Elemente erklärt werden können – d.h. ohne dafür Parallelverarbeitung längerer lokaler Strings annehmen zu müssen. Wie oben erwähnt ist ein zentrales Problem dieses Ansatzes, dass er LSK-Effekte nur als Unterschied von ambigen gegenüber nicht-ambigen lexikalischen Elementen vorhersagen kann. Demgegenüber wurden aber auch LSK-Effekte zwischen Sätzen mit nicht-ambigem Verb festgestellt, die durch ein eingeschobenes Adverb manipuliert wurden (vgl. Konieczny, Müller, Baumann et al., 2009; Weldle, Konieczny, Wolfer et al., 2010).

**Effekte des Gedächtnisabrufs** Einen anderen Ansatz stellen R. L. Lewis et al. (2005); R. L. Lewis und Vasishth (2006) und darauf aufbauend van Dyke (2007) vor. Die Erklärungsgrundlage sind ebenfalls die Verarbeitungseigenschaften für lexikalische Elemente im Satz, wobei aber die Erklärungen über Prozesse des Abrufs und der Übereinstimmung von Informationen aus dem Arbeitsgedächtnis erfolgen. Die Grundfrage ist: *„How are linguistic relations established in sentence processing – exactly what are the working memory processes that bring prior linguistic material into contact with present material, and what are the constraints on those processes?“* (R. L. Lewis et al., 2005, S. 376, Hervorhebung im Original). In ihrer Theorie erfolgt Satzverarbeitung als Abfolge spezialisierter, assoziativer Gedächtnisabruf-Prozesse, die durch Aktivierungsausbreitung und ähnlichkeitsbasierte Interferenzen im Arbeitsgedächtnis reguliert werden. Die hierfür angenommenen Prinzipien wurden in der kognitiven Architektur ACT-R (J. R. Anderson & Lebiere, 1998; J. R. Anderson, 2005) formalisiert und als generatives Modell implementiert. Als zentrale Verarbeitungsprinzipien und -eigenschaften des Arbeitsgedächtnisses, die der Sprachverarbeitung zugrunde liegen, werden identifiziert: (a) Ein limitierter Aufmerksamkeitsfokus, (b) schneller, inhaltsadressierter Zugriff auf Item-Informationen, aber nicht auf sequenzielle Informationen, (c) ähnlichkeitsbasierte Interferenzen beim Abruf von Informationen, (d) ähnlichkeitsbasierte Interferenzen bei der Enkodierung von Informationen, und (e) Aktivierungsausbreitung in Abhängigkeit von Aktivierungszersplitterung und bisherigen Abrufen. Demzufolge sind die zentralen Beschränkungen bei der Sprachverarbeitung nicht Beschränkungen der Speicherkapazität, sondern Beschränkungen des Arbeitsgedächtnisses, verschiedene ähnliche lexikalische Elemente zu diskriminieren (vgl. R. L. Lewis & Vasishth, 2006).

In diesem Cue-Based Parsing-Ansatz erklärt van Dyke (2007) als Ursache der von Tabor et al. (2004) gezeigten LSK-Effekte Interferenzen bei Abruf auf dem Verb durch konkurrierende Nominalphrasen. Diese Interferenzen werden im Cue-Based Parsing sowohl von syntaktischen als auch von semantischen Eigenschaften der konkurrierenden Elemente ausgelöst und können sich in diesem Ansatz auch ohne LSK auf die Verarbeitung auswirken. LSK sind also nur ein Effekt von NP-Interferenzen, der in die Klasse der Interferenz-Effekte subsumiert werden kann. In ihren Experimenten untersucht van Dyke dies mit Stimuli,

<sup>11</sup> Allerdings wird neben der korpusgestützten Bestimmung der Wahrscheinlichkeitsverteilungen der kontextunabhängigen Kategorien für einzelne lexikalische Elemente nicht ganz klar, wie die top-down Wahrscheinlichkeiten bestimmt werden.

in denen verschiedene Interferenz-Quellen auf dem Verbkomplex *was complaining* um ihre Anbindung konkurrieren, während korrekt die NP *the resident* angebunden werden müsste. In den Beispielen 14b und 14d ist die Interferenz durch die belebte NP *the neighbor* semantischer Natur, in den Beispielen 14c und 14d ist sie syntaktischer Natur, da die interferierende NP *the warehouse/the neighbor* die syntaktische Funktion eines Satzsubjekts erfüllt. Unabhängig davon, ob der jeweilige Satz eine LSK enthält (wie in den Bsp. 14a und 14b), sollten die Sätze, die eine der jeweiligen Interferenz-Quellen enthalten, stärkere Effekte auf dem kritischen Verbkomplex (*was complaining*) aufweisen, als die Sätze ohne die jeweilige Interferenz-Quelle.

- (14) The worker was surprised that the resident...
- a. who was living near the dangerous warehouse was complaining  
(keine Interferenz)
  - b. who was living near the dangerous neighbor was complaining  
(semantische Interferenz)
  - c. who said that the warehouse was dangerous was complaining  
(syntaktische Interferenz)
  - d. who said that the neighbor was dangerous was complaining  
(syntaktische und semantische Interferenz)
- ...about the investigation.

Diese Hypothese wurde mit Lesezeit-Experimenten überprüft (Self-Paced Reading mit folgender Verständnis-Befragung, Blickbewegungsaufzeichnung während des Lesens). Die Resultate zur Akkuratheit des Satzverstehens<sup>12</sup> weisen für beide Interferenz-Arten Haupteffekte auf. Allerdings findet sich der Effekt semantischer Interferenz nur für die Bedingungen 14a und 14b – was eher den Vorhersagen von LSK-Ansätzen entspricht. Auch die Latenzen der Lesezeitmessung auf der kritischen und der spill-over Region sprechen eher für einen LSK-Effekt für Bedingung 14b entsprechend den Befunden von Tabor et al. (2004). Dem hält van Dyke (2007) einen späten Effekt erhöhter Lesezeiten auf dem letzten Wort (*investigation*) in Bedingung 14d entgegen, der einem zusätzlichen Effekt semantischer Interferenz ohne vorhandene LSK entsprechen soll. In den Blickbewegungsaufzeichnungen während des Lesens entsprechen die Ergebnisse hingegen den Annahmen des Cue-Based-Parsing-Ansatzes mit Effekten beider Interferenz-Arten über alle Bedingungen (also auch semantische Interferenz-Effekte ohne LSK, Bed. 14d). Allerdings ist es schwierig, aus den gegebenen Daten eine klare Distinktion verschiedener Ansätze abzuleiten, denn abgesehen davon, dass die Effektstruktur alles andere als klar und übersichtlich ist, scheinen die Ergebnisse ansonsten mit den LSK-zentrierten Ansätzen kompatibel zu sein.

**Ressourcen-Allokation/Abgleich statistischer Erwartungen** Neben Theorien, die an Arbeitsgedächtnis-Ressourcen und Informationsabruf orientiert sind, wurden mittlerweile

---

<sup>12</sup>Zugehörige Aufgabe der VTn war es, per Selbsteinschätzung zu bewerten, ob sie den jeweiligen Satz verstanden haben. Die Akkuratheit gibt die Rate der Sätze wieder, die als verstanden bewerteten wurden. Dieses Maß als Grundlage für die Beurteilung des Satzverstehens zu verwenden ist allerdings höchst problematisch, von objektiver Messung kann hier kaum gesprochen werden. Eine bessere Alternative wäre es gewesen, das Verstehen direkt nachzuprüfen (z.B. durch Verständnisfragen).

mehrere Ansätze vorgeschlagen, die der Forschungslinie probabilistischer Sprachverarbeitungstheorien entstammen. Ein früherer Ansatz, der sich auch auf LSK bezieht, wurde von R. Levy (2008) entwickelt. Seine Theorie der constraintbasierten Ressourcen-Allokation auf Grundlage paralleler probabilistischer Auflösung von Ambiguitäten lässt sich formal als Äquivalent zur Surprisal-Theorie (Hale, 2001) abbilden. Im Sprachverarbeitungssystem werden alle zu einem partiellen Input konsistenten strukturellen Analysen parallel erfasst und gerankt. Verarbeitungskomplexität entspricht dem Umfang der Neuverteilung von Ressourcen durch ein Wort in diesem parallelen Ranking – d.h. welcher Umfang an Neustrukturierung des Rankings alternativer Strukturen wird durch ein neues Wort benötigt. Je umfangreicher diese Reallokation von Ressourcen ausfällt, desto größer sind die erwarteten Verarbeitungsschwierigkeiten bzw. Lesezeiten. Während dieser Ansatz eine große Abdeckung verschiedenster psycholinguistischer Phänomene aufweist, bereiten ihm LSK-Effekte Schwierigkeiten. Um diese Effekte zu erfassen, müsste in dieses vollständig inkrementelle Modell ein Maß der Unsicherheit über die vorausgegangenen Wörter eingeführt werden. Die Unsicherheit darüber, welche Wörter tatsächlich gehört wurden, könnte zur Annahme nicht-gehörter Wörter führen, mit denen LSK auch global kohärent werden und als mögliche Strukturen in das parallele Ranking eingehen können.<sup>13</sup>

Bicknell und Levy (2009) beschäftigen sich speziell mit dieser Lücke. Sie schlagen als Erklärungsansatz für LSK-Effekte in Entsprechung zu R. Levy (2008) den Aufwand zum Update von Strukturannahmen in einem probabilistischen Earley Parser (Stolcke, 1995) vor. Im Zentrum steht der Umfang der Umstrukturierung der kontextunabhängigen beliefs (probabilistische Annahmen) über die syntaktische Struktur einer begrenzten Sequenz von Wörtern zu den beliefs über die syntaktische Struktur im globalen Kontext des Satzes. Ähnlich wie im Ansatz von Gibson (2006), aber eben nicht begrenzt auf einzelne lexikalische Elemente, geht es also gewissermaßen darum, kontextunabhängige Bottom-up-Informationen und kontextabhängige Top-down-Informationen zu integrieren. Je größer der Unterschied zwischen diesen ist, desto größer sind die Verarbeitungsschwierigkeiten. Da der Unterschied für Sätze mit LSK entsprechend groß ausfällt, kommt es zu den entsprechenden Effekten erhöhter Lesezeiten, wie die Autoren mit ihrem Parsing-Modell für die Stimuli von Tabor et al. (2004) demonstrieren. Das scheinbar irrationale Verhalten für LSK wird damit zum rationalen Effekt des Abgleichs statistischer Erwartungen aus verschiedenen Informationsquellen.

Zu erwähnen wäre noch der Ansatz von Hale (2011), der eine breit angelegte Rahmentheorie der Satzverarbeitung auf Basis klassischer rationaler Verarbeitungsprinzipien konzipiert. Die Beschränkung und zugleich die Rationalität dieses Parsing-Modells liegt in der Prämisse des Ziels, inkrementell auf der Basis einer probabilistisch informierten Suche nach syntaktischen Alternativen für Strukturierung und Anbindung möglichst schnell eine möglichst gute syntaktische Analyse für einen Satz zu finden. LSK-Effekte werden in

<sup>13</sup>Dieser Ansatz trägt ein Problem der kombinatorischen Explosion möglicher Alternativinterpretationen in sich: Wenn nicht nur die parallelen Interpretationen für die gegebenen Wörter, sondern auch die zusätzlichen Interpretationen aufgrund der auf jeder Position möglichen alternativen Wörter relevant werden, wird die Zahl zu berücksichtigender Interpretationen unüberschaubar groß. Zum einen ist unklar, wie ein beschränktes natürliches System eine solche Informationsmenge verarbeiten soll. Zum anderen lässt es sich methodisch schwerlich bewerkstelligen, Experimente bzw. Stimulusmaterialien zu erstellen, in denen Alternativinterpretationen durch unsichere Wörter kontrolliert werden. Damit kann die Hypothese von R. Levy (2008) nicht sinnvoll mit psycholinguistischen Experimenten überprüft werden.

diesem Ansatz als serieller Wettbewerb um Verarbeitungsressourcen bzw. um benötigte Verarbeitungszeit erklärt, der bei der probabilistisch informierten Suche nach syntaktischen Alternativen entsteht. In Sätzen mit lokal ambigen Strukturen (vgl. Tabor et al., 2004; Konieczny & Müller, 2007) müssen dabei im Verhältnis wesentlich mehr Suchzustände bewältigt werden.

### 5.3. Anaphernaufflösung

Die Anapher bildet zusammen mit der Deixis eine Klasse referenzieller Ausdrücke. Während deiktische Ausdrücke<sup>14</sup> direkt auf Aspekte der Äußerungssituation referieren, also direkt über den situationalen Kontext interpretiert werden (exophorische Ausdrücke), verweisen Anaphern nicht direkt auf einen Referenten in der Welt, sondern indirekt über ein anderes Wort im Diskurs<sup>15</sup> (endophorische Ausdrücke), das Antezedens der Anapher, und koreferieren mit diesem auf einen Referenten (vgl. z.B. Meibauer et al., 2002; Lehmann, 2007). Das bedeutet, dass auch Anaphern letzten Endes auf einen Referenten in der Welt verweisen, zur Interpretation aber ihre Identifikation über den linguistischen Koreferenten benötigen. Zu den Anaphern gehören verschiedene pronominale Formen, unter anderem die Personalpronomen (z.B. *er, sie, es*), das Reflexivpronomen *sich*, und die Relativpronomen (z.B. *der, die*). Neben den pronominalen gibt es nicht-pronominale Formen, beispielsweise Nominalphrasen, die den zuvor genannten Referenten wieder aufgreifen. Im Folgenden wollen wir uns speziell mit pronominalen Anaphern beschäftigen.

Für die psycholinguistische Forschung stellt sich die Frage, wie solche anaphorischen Ausdrücke an ein Antezedens gebunden werden, d.h. wie das korrekte Antezedens identifiziert und dadurch die Anapher in ihrer Koreferenz interpretiert wird: Welche Constraints regeln während der Verarbeitung eines anaphorischen Ausdrucks in einem Satz, welches der genannten Argumente das Antezedens sein kann, eventuell, welches als Antezedens ausgeschlossen ist, eventuell gar, welches das Antezedens sein muss – und welcher Natur sind diese Constraints (semantisch, pragmatisch, strukturell, syntaktisch, grammatisch)? Welcher Natur sind die kognitiven Mechanismen, über die die Filterung und Auswahl des Antezedens geleistet wird, einschließlich ihres zeitlichen Verlaufs und ihrer Interaktionen mit anderen kognitiven Mechanismen und Eigenschaften (etwa Ressourcen wie das Arbeitsgedächtnis, oder Aspekte der linguistischen Erfahrung)?

Unter den theoretischen Ansätzen lassen sich zwei Arten von Mechanismen unterscheiden: die, die in irgendeiner Form beschränkend wirken, d.h. zwingende Vorgaben machen und distinktiv sind (diese werden im weiteren Verlauf *Bindungs-Constraints* genannt), und die, die in irgendeiner Form Vorschläge oder Gewichtungen der Wahrscheinlichkeit eines Koreferenten machen, d.h. Präferenzen widerspiegeln (diese werden im Folgenden als *Bindungspräferenzen* bezeichnet). Mitkov (1999) gibt allerdings zu bedenken, dass eine

---

<sup>14</sup>Zu diesen zählen unter anderem Personaldeixis (*ich, du*), Objektdeixis (*dieser, jenes*) und Temporaldeixis (*jetzt, morgen*).

<sup>15</sup>Bei Anaphern richtet sich der Verweis auf einen bereits genannten Diskurs-Referenten, bei Kataphern auf einen erst noch folgenden Diskurs-Referenten. Im Sinne einer textlinguistischen Definition sind Anaphern eine Form der Kohäsion (bzw. ein kohäsionsstiftendes Element), die auf eine vorausgegangene sprachliche Einheit zurück verweist (vgl. Halliday & Hasan, 1976).

solche Unterscheidung zumindest fragwürdig ist: Statt der Annahme harter Constraints wäre es möglicherweise angemessener, alle Mechanismen im Sinne von Präferenzen zu verstehen, die sich in stärkerem oder schwächerem Maße beschränkend auswirken. Tabelle 5.1 bietet einen Überblick zu den gängigsten Ansätzen (nach Mitkov, 1999).

CONSTRAINTS	THEORETISCHER ANSATZ	SATZBEISPIEL
Bindungs-Constraints		
Grammatische Constraints	Genus und Numerus von Antezedens und Anapher müssen übereinstimmen. Bei mehreren Kandidaten wird nach grammatischer Übereinstimmung gefiltert.	Sandra <sub>i</sub> erzählte Philip <sub>j</sub> und seinen Freunden <sub>k</sub> , dass sie <sub>i</sub> verliebt ist.
Strukturelle Constraints	Die Beschränkung möglicher Antezedenzen erfolgt nach strukturellen Prinzipien, wie etwa in der Government&Binding-Theorie über C-Kommando-Strukturen anhand von Prinzipien beschrieben werden (siehe unten).	Daniel <sub>i</sub> erzählt Sascha <sub>j</sub> über sich <sub>i</sub> .
Semantische Constraints	Konsistenz semantischer Merkmale (der Anapher durch ihren Kontext vorgegeben), zum Beispiel Belebtheit, Verwendung als Instrument, mögliche Handlungs- oder Verwendungsweise.	Peter entfernte den USB-Stick <sub>i</sub> aus dem Computer <sub>j</sub> , und schaltete ihn <sub>i</sub> danach aus.
Bindungspräferenzen		
Centering Theory	Bevorzugung des salientesten Antezedens, d.h. des Objekts, auf dem innerhalb des Diskurses durch Hervorhebung der Fokus liegt.	Franziska war gestern auf Shopping-Tour und kaufte sich eine wunderschöne Teetasse <sub>i</sub> . Sie zeigte die Tasse <sub>i</sub> allen ihren Frunden. Als sie allerdings nach Hause zurückgekehrt war, stellte sie die Teetasse auf eine Kuchenplatte <sub>j</sub> und zerbrach sie <sub>i</sub> .
Syntaktischer Parallelismus	Bevorzugung des Antezedens mit gleicher syntaktischer Funktion.	Der Programmierer <sub>i</sub> kombinierte beim letzten Mal R <sub>j</sub> mit Python <sub>k</sub> , beim nächsten Mal kombiniert er <sub>i</sub> es <sub>j</sub> mit Lisp.
Semantischer Parallelismus	Bevorzugung des Antezedens mit gleicher semantischer (bzw. thematischer) Rolle, zum Beispiel Empfänger einer Handlung.	Lars <sub>i</sub> schenkte Sascha <sub>j</sub> eine Mother's Finest-CD, und Daniel <sub>k</sub> schenkte ihm <sub>j</sub> die neueste 24-Staffel.

Tabelle 5.1.: Theoretische Ansätze zur Erklärung von Constraints und Präferenzen, die der Bindung von Anaphern an ihr Antezedens zugrunde liegen.

### 5.3.1. Strukturelle Ansätze: Bindungsdomänen

In nicht-ambigen Kontexten erfolgt die Auflösung von Anaphern anhand einer Kombination aus grammatikalischen und strukturellen Merkmalen. Morphosyntaktische Merkmale – die Übereinstimmung von Genus und Numerus zu einem Antezedens – schränken die möglichen Antezedenzen auf Wortebene ein. Strukturelle Merkmale ermöglichen Constraints auf syntaktischer Ebene und fungieren als weitere Filterbedingungen der Auswahl. Strukturelle Constraints werden in verschiedenen Grammatiktheorien etwa im Sinne syntaktischer Kommando-Relationen oder sonstiger struktureller Relationen beschrieben und formalisiert, so etwa in der Government&Binding-Theorie (Chomsky, 1981) als C-Kommando, in der Head-Driven-Phrase-Structure-Grammar (Pollard & Sag, 1994) als O-Kommando,

oder in der Lexical Functional Grammar (vgl. R. Kaplan & Bresnan, 1982) als F-Struktur-Konfiguration.

Unser Hauptaugenmerk liegt im folgenden Abschnitt auf den rein strukturell formulierten Formalisierungen der Constraints für die Anaphernauffösung, die mit der Bindungstheorie (Chomsky, 1981) entwickelt wurden (vgl. z.B. Fanselow & Felix, 1987; Dürscheid, 1991; Büring, 2005).<sup>16</sup> In dieser werden die möglichen und zwingenden Koreferenzen eines anaphorischen Ausdrucks durch strukturelle Prinzipien bedingt, die anhand von C-Kommando-Relationen beschrieben werden können. Mit C-Kommando-Relationen werden syntaktische Beziehungen zwischen Konstituenten erfasst. Dabei c-kommandiert ein Knoten<sup>17</sup> A einen Knoten B, wenn A und B vom gleichen Knoten dominiert werden (d.h. beide entstammen der gleichen Verzweigung), zugleich aber A nicht von B und umgekehrt B nicht von A dominiert wird. Wenn A B c-kommandiert, dann c-kommandiert A auch die Folgeknoten von B. Durch diese strukturelle Beziehung wird auch die Bindung zweier Knoten beschrieben: A bindet B, wenn A B vorausgeht, A B c-kommandiert und A und B koindiziert (also bezüglich ihrer Referenz identisch markiert) sind. Erläuternde Beispiele folgen in den nächsten Abschnitten.

Auf Grundlage dieser Relation zwischen syntaktischen Konstituenten können allgemeine grammatische Prinzipien bzw. Beschränkungen formuliert werden. Im Zentrum der Beschränkungen der Koreferenz syntaktischer Konstituenten (d.h. Beschränkung, welche Wörter oder Phrasen auf den gleichen gemeinsamen Referenten verweisen können) stehen die Prinzipien A, B, und C. Diese beschreiben verschiedene strukturelle Voraussetzungen bzw. Bindungsregularitäten, die für die Koreferenz verschiedener referenzieller Ausdrücke – reflexive Pronomen, nicht-reflexive Pronomen und sogenannte R-Ausdrücke (für referential expressions, d.h. Nominalphrasen, Eigennamen) – gegeben sein müssen. In folgender Liste werden die Bindungsprinzipien A, B und C aufgeführt und in ihrer Relevanz für verschiedene Realisierungen referenzieller Formen in Relation zur zugrundeliegenden C-Kommando-Struktur gestellt:

**Prinzip A** beschreibt die strukturellen Voraussetzungen der Koreferenz für reflexive Pronomen, zum Beispiel *sich*, *dich*, *mich*. Reflexive Pronomen müssen innerhalb der Bindungsdomäne des Antezedens gebunden sein, d.h. sie müssen durch das Antezedens c-kommandiert und mit ihm koindiziert sein. Die Bindung muss lokal im gegenwärtigen Satz erfolgen.

**Prinzip B** beschreibt die strukturellen Voraussetzungen der Koreferenz für nicht-reflexive Pronomen, zum Beispiel Personalpronomen wie *sie*, *ihn*, *es*. Nicht-reflexive Pronomen dürfen nicht innerhalb der Bindungsdomäne des Antezedens gebunden bzw. müssen innerhalb dieser frei sein, d.h. sie dürfen nicht durch das Antezedens c-

---

<sup>16</sup>Es gibt verschiedene Weiterentwicklungen der Bindungstheorie. So diskutiert zum Beispiel S. Fischer (2004) Erweiterungen der bestehenden Theorie im Sinne einer optimalen Bindungstheorie (im Sinne einer Bindungstheorie in der Optimality Theory), die Phänomene wie Abweichungen bei Fernabhängigkeiten, Verletzung der Komplementaritätsannahme, Unterschiede zwischen verschiedenen Anaphern-Typen und Bindung von R-Expressions erfasst. Für unsere Belange sind aber zunächst nur die Grundprinzipien relevant.

<sup>17</sup>Mit Knoten bezeichnen wir ein terminales oder nicht-terminales Element in einem Phrasenstrukturschema, etwa in der Tradition der Generativen Grammatik der X-Bar-Struktur.

kommandiert und koindiziert sein. Die Bindung beschränkt sich lokal auf den gegenwärtigen Satz.

**Prinzip C** beschreibt die strukturellen Voraussetzungen der Referenz für referenzielle nominale Ausdrücke, also NPs einschließlich Eigennamen (z.B. *Peter*, *der Polizist*). Diese dürfen weder innerhalb noch außerhalb der Bindungsdomäne eines vorausgegangenen referenziellen Ausdrucks gebunden sein, d.h. sie dürfen nicht koreferieren.

Reflexive Pronomen wie Reflexivpronomen oder Rezipropronomen müssen also innerhalb ihrer C-Kommando-Struktur gebunden sein, während nicht-reflexive Pronomen wie Personal- oder Possessivpronomen nicht innerhalb ihrer C-Kommando-Struktur gebunden sein dürfen. In der Bindungstheorie wird anhand dieses Unterschieds auch die terminologische Unterscheidung zwischen Anaphern und Pronomen begründet. Reflexivpronomen und Personalpronomen weisen also komplementäre Bindungsdomänen auf (vgl. auch Runner, Sussman & Tanenhaus, 2003; R. Frank et al., 2005; Kaiser et al., 2009).

**Beispiele struktureller Beschränkungen** Die nachfolgenden Abbildungen (Abb. 5.1, 5.2 und 5.3) illustrieren die Prinzipien A und B anhand von Beispielsätzen in Phrasenstrukturbäumen.

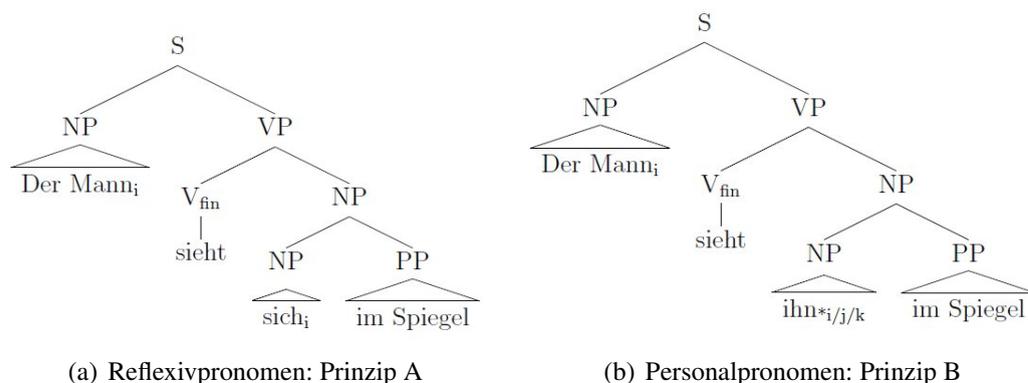


Abbildung 5.1.: Links: Strukturbaum zum Satz *Der Mann sieht sich im Spiegel*. Für Reflexivpronomen wird die strukturelle Beschränkung der Koreferenz durch Prinzip A beschrieben. Im Vergleich dazu rechts: Strukturbaum zum Satz *Der Mann sieht ihn im Spiegel*. mit Personalpronomen. Hier beschreibt Prinzip B die strukturelle Beschränkung der Koreferenz.

Im Satz *Der Mann sieht sich im Spiegel*. (Abb. 5.1(a)) bindet der Knoten NP *der Mann* den Knoten N *sich*. Der ihnen gemeinsame Mutterknoten S dominiert beide Knoten und der Knoten NP *der Mann* geht dem Knoten N *sich* voraus – der Knoten NP *der Mann* c-kommandiert also den Knoten N *sich* – und die beiden Knoten sind koindiziert. Folglich muss das Reflexivpronomen *sich* nach Prinzip A in diesem strukturellen Kontext mit der NP *der Mann* koreferieren. Im Satz *Der Mann sieht ihn im Spiegel*. (Abb. 5.1(b)) dominiert der gemeinsame Mutterknoten S die beiden Knoten NP *der Mann* und N *ihn* und der Knoten NP *der Mann* geht dem Knoten N *ihn* voran. Folglich c-kommandiert der Knoten NP *der Mann* den Knoten N *ihn*. Da das Personalpronomen *ihn* nach Prinzip B

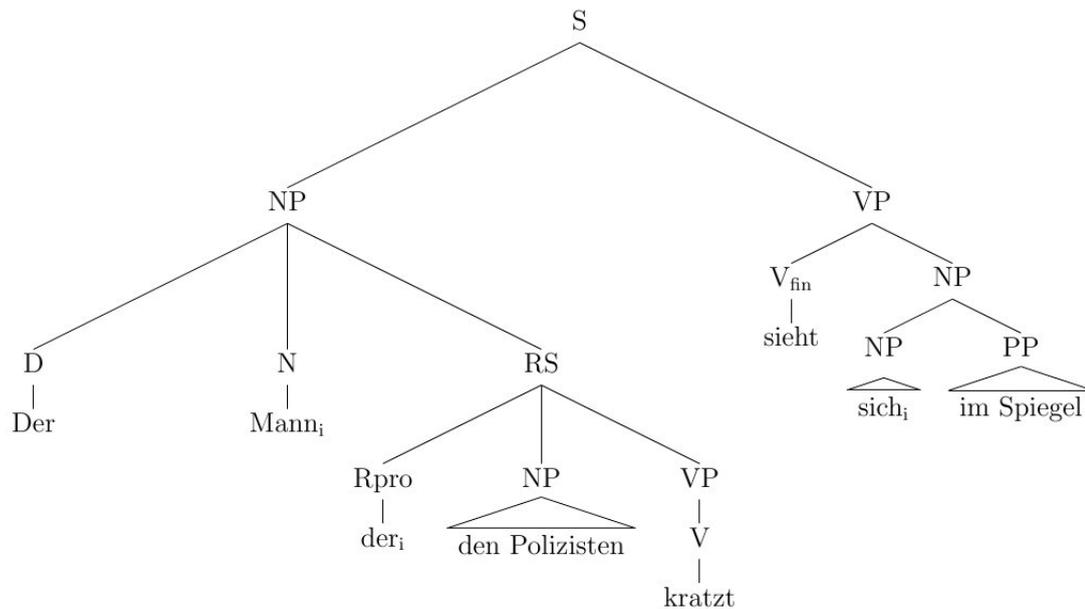


Abbildung 5.2.: Strukturbaum zum Satz *Der Mann, der den Polizisten kratzt, sieht sich im Spiegel.* Strukturelle Beschränkung der Koreferenz durch Prinzip A.

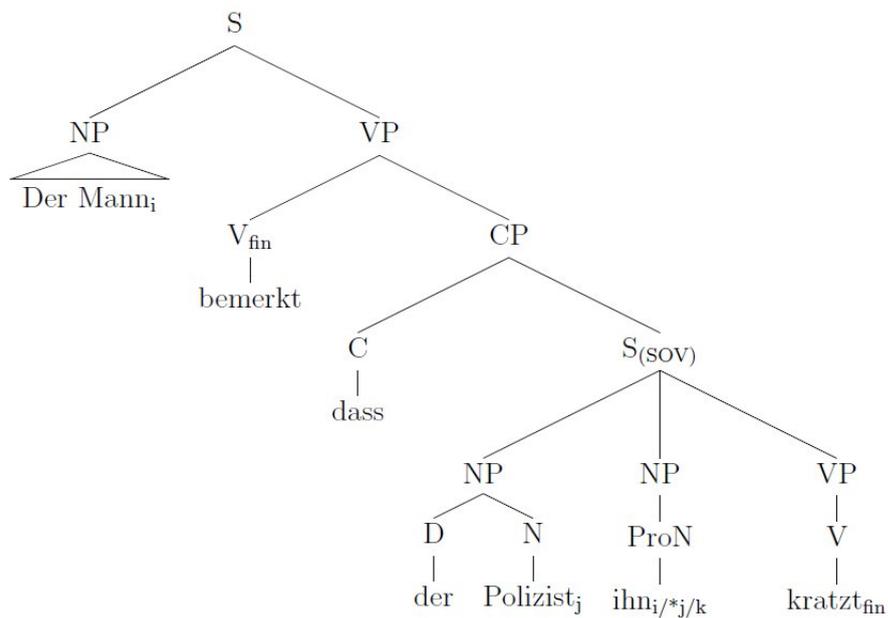


Abbildung 5.3.: Strukturbaum zum Satz *Der Mann bemerkt, dass der Polizist ihn kratzt.* Strukturelle Beschränkung der Koreferenz durch Prinzip B.

nicht gebunden sein darf, ist der Knoten NP *der Mann* in diesem strukturellen Kontext nicht als Koreferent des Knotens N *ihn* zugänglich. Stattdessen kann der Knoten N *ihn* an jedes beliebige andere Antezedens gebunden werden. Prinzip B beschreibt also nicht einen determinierenden strukturellen Zusammenhang, sondern schließt ein unzugängliches Antezedens als Koreferenten aus.

Mit den folgenden beiden Sätzen erläutern wir die Prinzipien A und B in komplexeren syntaktischen Kontexten. Im Satz *Der Mann, der den Polizisten kratzt, sieht sich im Spiegel.* (Abb. 5.2) bindet der Knoten NP *der Mann* den Knoten N *sich* – denn auch hier dominiert der gemeinsame Mutterknoten S beide Knoten. Unabhängig vom intervenierenden Relativsatz, der an die NP *der Mann* angebunden ist, c-kommandiert der Knoten NP *der Mann* also den Knoten N *sich* und die beiden Knoten sind koindiziert. Folglich muss das Reflexivpronomen *sich* nach Prinzip A auch in diesem strukturellen Kontext mit der NP *der Mann* koreferieren. Im Satz *Der Mann bemerkt, dass der Polizist ihn kratzt.* (Abb. 5.3) dominiert der gemeinsame Mutterknoten S die beiden Knoten NP *der Mann* und N *ihn* und der Knoten NP *der Mann* geht dem Knoten N *ihn* voran. Dementsprechend müsste nach Prinzip B die NP *der Mann* als Koreferent für das Personalpronomen *ihn* unzugänglich sein. Allerdings gilt für die Prinzipien A und B wie oben beschrieben ein Lokalisierungsprinzip: Um das Bindungsverhältnis zu bestimmen, müssen sich der c-kommandierende und der c-kommandierte Knoten innerhalb eines Satzes befinden. Der Knoten NP *der Mann* befindet sich aber im Matrixsatz, der Knoten N *ihn* hingegen im subordinierten Komplementsatz. Folglich ist das Personalpronomen *ihn* nicht in der Bindungsdomäne des Knotens NP *der Mann* gebunden und kann nach Prinzip B in diesem strukturellen Kontext mit dieser NP *der Mann* koreferieren – oder aber mit einem beliebigen anderen Antezedens.

### 5.3.2. Online-Anaphernauflösung: Empirische Befunde

Lassen sich die bindungstheoretischen Annahmen der Government&Binding-Theorie direkt auf die Online-Verarbeitung übertragen? Beschreiben die hier angenommenen Prinzipien tatsächlich die Beschränkungen und Präferenzen, die wir in empirischen Untersuchungen finden? Inwiefern bieten die strukturellen Prinzipien Anhaltspunkte für die Prozesse und Mechanismen, die für die Auflösung der anaphorischen Referenz verantwortlich sind? Und verläuft die Auflösung zwingend nach strukturellen Prinzipien, oder gibt es Einflüsse und Interferenzen auf die (semantisch-referenzielle) Verarbeitung anaphorischer Ausdrücke? In der Bindungstheorie selbst werden keine Annahmen zur eigentlichen Verarbeitung getroffen. Dennoch werden die hier postulierten Prinzipien mitunter als harte Constraints der Verarbeitung angenommen. In verschiedenen Studien wird eine strukturell determinierte Verarbeitung für Reflexiv- und Personalpronomen mit komplementärer Verteilung der für die Koreferenz zugänglichen Antezedenzen unterstellt.

In der jüngeren Forschung werden verschiedene Faktoren hinsichtlich ihres Einflusses und des Zeitpunkts ihres Einflusses auf die Verarbeitung von Anaphern empirisch untersucht, wobei auch die Überprüfung der Annahme struktureller Bindungs-Constraints zur Disposition steht. Dabei wird der zeitliche Verlauf der Verarbeitung bis zur tatsächlichen referenziellen Interpretation und die Gewichtung und Reihenfolge der Berücksichtigung verschiedener Informationsquellen (syntaktisch-strukturelle, semantisch-pragmatische und vom Diskurskontext bestimmte Verarbeitung) sowie Quellen von Interferenzen thematisiert.

Vor allem die Annahme der Komplementarität der Bindungsdomänen für Reflexiv- und Personalpronomen wurde in den letzten Jahren kritisch hinterfragt.<sup>18</sup>

Die meisten empirischen Befunde weisen auf die Sensitivität der Verarbeitung für strukturelle Faktoren entsprechend der Bindungstheorie hin. Allerdings gibt es in beinahe allen Studien auch Hinweise auf Einflüsse anderer Faktoren und inzwischen finden sich in speziellen syntaktischen Kontexten vermehrt Evidenzen gegen das Primat struktureller Faktoren der Anaphernauffösung. Für diese Befunde wurden verschiedene Erklärungsansätze entwickelt, die der strukturell bedingten Verarbeitung verschiedene Rollen zuweisen. Die folgenden Abschnitte geben einen Überblick über verschiedene Ansätze zu Faktoren und zeitlicher Koordination der Anaphernverarbeitung (eine ausführlichere Diskussion findet sich z.B. in Runner et al., 2006). Einige Aspekte, die wir in den späteren Kapiteln wieder aufgreifen, werden näher erläutert.

**Strukturelle Prinzipien als initialer Filter** In der Early-Filter-Hypothese (Nicol & Swinney, 1989) dienen strukturelle Prinzipien als initialer Filter zur frühen Auswahl möglicher Antezedenzen und zum Ausschluss strukturell inkompatibler Antezedenzen. Erst in der weiteren Verarbeitung kommen semantische, pragmatische und sonstige inhaltsbezogene Faktoren zum Einsatz, die die strukturellen Beschränkungen aber nicht verletzen können. Befunde stammen aus Studien mit modalitätsübergreifendem lexikalischem Priming (Abfrage der Antezedenzen auf verschiedenen Satzpositionen mit Lexical Decision-Aufgabe) für auditiv präsentierte Komplement- und Relativsätze. Dazu wurden auf der Satzposition nach der Anapher (im Beispiel markiert mit Asterisk, vgl. Bsp. 15) die potentiellen Antezedenzen (*boxer*, *skier* und *doctor*) als Prime dargeboten. Für diese zeigt sich ein schnellerer bzw. vereinfachter Abruf, wenn das Antezedens kompatibel zu den strukturellen Beschränkungen ist.

- (15) a. The boxer<sub>i</sub> told the skier<sub>j</sub> that the doctor<sub>k</sub> for the team would blame himself<sub>k</sub> \*  
for the recent injury.
- b. The boxer<sub>i</sub> told the skier<sub>j</sub> that the doctor<sub>k</sub> for the team would blame him<sub>i/j</sub> \*  
for the recent injury.

**Multiple interaktive Constraints** Mit der Hypothese multipler interaktiver Constraints nehmen Badecker und Straub (2002) hingegen an, dass strukturelle und inhaltliche Faktoren (Morphosyntax, Semantik, Diskursfokus, etc.) und Beschränkungen gleichzeitig verarbeitet werden und bei der Verarbeitung interagieren. Dies schließen sie aus Befunden einer Serie von Self-Paced Reading-Studien mit Komplementsätzen, in denen sich Interferenz-Effekte strukturell unzugänglicher Antezedenzen zeigen, wenn mehrere Antezedenzen mit gleichen semantischen bzw. morphosyntaktischen Merkmalen unabhängig

---

<sup>18</sup>Weitere Fragen betreffen unter anderem die Rolle der Anaphernauffösung bei global ambiger syntaktischer Anbindung und Differenzen einzelsprachlicher Präferenzen. So können über ambige Relativpronomen Anbindungsambiguitäten für Relativsätze entstehen, die es ermöglichen, in sprachübergreifenden Studien experimentell einzelsprachliche syntaktische Anbindungspräferenzen zu bestimmen (vgl. z.B. Hemforth, Konieczny & Scheepers, 2000; Hemforth & Konieczny, 2002). Diese Aspekte sind aber für die weiteren Fragestellungen irrelevant.

ihrer strukturellen Zugänglichkeit um die Bindung konkurrieren. So führt das Vorhandensein zweier Antezedenzen, die hinsichtlich des Genus kongruent zum Personalpronomen oder zum Reflexivpronomen sind (Bsp. 16b und 16d), zu längeren Lesezeiten auf den Folgepositionen nach der Anapher, als wenn nur ein kongruentes Antezedens gegeben ist (Bsp. 16a und 16c). Das trifft selbst dann ein, wenn das jeweils zweite konkurrierende Antezedens nach strukturellen Beschränkungen nicht zugänglich sein sollte.

- (16) a. John<sub>i</sub> thought that Beth<sub>j</sub> owed him<sub>i/k</sub> another opportunity to solve the problem.  
*Personalpronomen, ein passendes Antezedens*
- b. John<sub>i</sub> thought that Bill<sub>j</sub> owed him<sub>i/k/\*j</sub> another opportunity to solve the problem.  
*Personalpronomen, multiple passende Antezedenzen*
- c. Jane<sub>i</sub> thought that Bill<sub>j</sub> owed himself<sub>j</sub> another opportunity to solve the problem.  
*Reflexivpronomen, ein passendes Antezedens*
- d. John<sub>i</sub> thought that Bill<sub>j</sub> owed himself<sub>j/\*i</sub> another opportunity to solve the problem.  
*Reflexivpronomen, multiple passende Antezedenzen*

**Mehrstufige Modelle** Sturt (2003) stellt dem parallel-interaktiven Ansatz einen zweistufigen Ansatz entgegen.<sup>19</sup> Er postuliert ein zweistufiges Verarbeitungsmodell: Auf der ersten Stufe wirken sich unmittelbar die strukturellen Beschränkungen aus. Erst auf der zweiten Stufe zeigen sich die Einflüsse anderer Faktoren, durch die die strukturellen Beschränkungen allerdings verletzt werden können. Diese Zuordnung wurde aus Studien mit Blickbewegungsmessung während des Lesens abgeleitet, die die Verarbeitung in einer höheren zeitlichen Auflösung abbilden und für deren Befunde eine Unterscheidbarkeit früher und später Effekte über verschiedene Lesezeitmaße angenommen wird (vgl. Kap. 5.2.2). Zur Untersuchung wurden Sätze verwendet, in denen sich eine NP mit unbestimmtem Genus befindet, für die aber ein stereotyper Genus vorherrscht (z.B. für das Nomen *surgeon* das Genus Maskulinum, vgl. Bsp. 17). Diese NP ist in den Experimentalsätzen strukturell für das Reflexivpronomen zugänglich, eine vorausgehende NP hingegen nicht. Das Reflexivpronomen kann hinsichtlich der Genus-Kongruenz dem Stereotyp der zugänglichen NP *surgeon* entsprechen oder nicht. Zugleich ist sie zur unzugänglichen NP (*He/Jonathan* oder *Her/Jennifer*) kongruent oder nicht-kongruent. In den frühen Maßen (First-Fixation- und First-Pass-Lesezeiten) findet sich auf der Anapher nur für das zugängliche Antezedens ein Mismatch-Effekt, wenn das Reflexivpronomen ein nicht-stereotypes Genus (*surgeon – herself*) vorgibt. Erst in den mit späten Prozessen assoziierten Maßen (hier: Second-Pass-Lesezeiten) findet sich ein Mismatch-Effekt (auf der vorfinalen Region *used syringe*) für das unzugängliche Antezedens. Frühe Prozesse scheinen also konform zu bindungstheoretischen strukturellen Constraints zu erfolgen, die aber in späteren Prozessen

<sup>19</sup>Ein ähnliches zweistufiges (bzw. vierstufiges) Modell wird von Kennison (2003) vorgeschlagen, die eine frühe Bindungsstufe mit einer vorausgehenden Auswahlstufe und eine späte Auflösungsstufe annimmt. Auf der initialen Stufe werden mögliche Antezedenzen identifiziert, wobei dies nicht alleine anhand struktureller Beschränkungen erfolgt. Diese greifen erst auf der Bindungsstufe, bei der die strukturell zugänglichen Antezedenzen an die Anapher gebunden werden. In der späten Auflösungsstufe schließlich wird die Auswahl getroffen. Eine zusätzliche Entscheidungsstufe greift als Reanalyseprozess, wenn die angenommenen Bindungen nicht in die semantische Struktur integriert werden können. In diesem Modell können sowohl in der frühen als auch in der späten Verarbeitung nicht-strukturelle Einflüsse wirken.

durch andere Faktoren (hier z.B. Fokus) verletzt werden können. Diese Effekte assoziiert Sturt eher mit Recovery-Strategien und Wrap-Up-Effekten. Wie ein Folgeexperiment (Self-Paced Reading mit anschließender Verständnisfrage) zeigt, können diese aber nachhaltig die Interpretation des Satzes beeinflussen.<sup>20</sup>

- (17) a. Jonathan<sub>i</sub> was pretty worried at the City Hospital. He<sub>i</sub> remembered that the surgeon<sub>j</sub> had pricked himself<sub>j</sub> with a used syringe needle. There should be an investigation soon.  
*Unzugänglich/kongruent – Zugänglich/stereotyp*
- b. Jennifer<sub>i</sub> was pretty worried at the City Hospital. She<sub>i</sub> remembered that the surgeon<sub>j</sub> had pricked himself<sub>j</sub> with a used syringe needle. There should be an investigation soon.  
*Unzugänglich/nicht-kongruent – Zugänglich/stereotyp*
- c. Jonathan<sub>i</sub> was pretty worried at the City Hospital. He<sub>i</sub> remembered that the surgeon<sub>j</sub> had pricked herself<sub>j</sub> with a used syringe needle. There should be an investigation soon.  
*Unzugänglich/kongruent – Zugänglich/nicht-stereotyp*
- d. Jennifer<sub>i</sub> was pretty worried at the City Hospital. She<sub>i</sub> remembered that the surgeon<sub>j</sub> had pricked herself<sub>j</sub> with a used syringe needle. There should be an investigation soon.  
*Unzugänglich/nicht-kongruent – Zugänglich/nicht-stereotyp*

**Semantisch-pragmatische Faktoren** Picture-Noun-Phrases (nachfolgend: PNP) sind Sätze mit einer Nominalphrase, deren Kopf ein repräsentierendes Nomen wie *Bild*, *Foto* oder auch *Geschichte* ist. Für diese ist einerseits bekannt, dass sie hinsichtlich struktureller Ansätze problematisch sind – zumindest im speziellen Fall der PNP ohne Possessor (Bsp. 18a), für die Verletzungen der Komplementaritätsannahme speziell für Reflexivpronomen nach Prinzip A gefunden wurden (vgl. z.B. Pollard & Sag, 1992; Keller & Asudeh, 2001; Asudeh & Keller, 2001). Dennoch wird im Allgemeinen für diese Struktur – etwa für PNP mit Possessor (Bsp. 18b) – angenommen, dass sie grundlegend strukturellen Prinzipien folgt, und nur PNP ohne Possessor eine Ausnahme darstellen. Um die hierfür gefundenen Verletzungen der bindungstheoretischen Prinzipien zu erfassen, wurden verschiedene Erklärungen vorgeschlagen, die spezielle Satzstrukturen als Ausnahmen klassifizieren, in denen die Bindung des Reflexivpronomens unter anderem nicht durch strukturelle, sondern durch pragmatische und Diskurs-Faktoren bestimmt werden.<sup>21</sup> Solche Reflexivpronomen

---

<sup>20</sup>Interessanterweise enthalten die Experimentalsätze in einem zweiten Experiment mit Blickbewegungsmessung während des Lesens aus der Studie von Sturt (2003) eine LSK. Der Satz lautet *Jonathan was pretty worried at the City Hospital. The surgeon who treated Jonathan had pricked himself with a used syringe needle. There should be an investigation soon.* und enthält die syntaktisch kohärente Subsequenz *Jonathan had pricked himself with a used syringe needle.* Entsprechend den Ergebnissen von Müller et al. (2009) würden wir in den Bedingungen mit kongruentem global unzugänglichem Antezedens eine geringere Lesezeit für die Anapher und die Folgeregionen in der LSK erwarten. Allerdings konnten in den für eine andere Fragestellung aufbereiteten und dargestellten Daten keine entsprechenden Effekte identifiziert werden. Zudem wird die LSK nach ihrem Abschluss nicht als inkompatible Fehlkonstruktion aufgelöst, wodurch der stärkere Effekt erhöhter Lesezeiten nicht gezeigt werden kann.

<sup>21</sup>Spezielle Theorien für dieses Problem der PNP ohne Possessor wurden im Sinne der Bindungstheorie mit der PRO-in-NP-Hypothese (z.B. Davies & Dubinsky, 2003) und in einem diskurstheoretischen

mit semantisch-pragmatisch aufgelöster Bindung werden als Logophore bezeichnet (vgl. Reinhart & Reuland, 1993).<sup>22</sup>

- (18) a. Peter<sub>i</sub> sieht ein Foto von ihm<sub>\*i/j</sub>/sich<sub>i/\*j</sub>.  
 b. Peter<sub>i</sub> sieht Lars' <sub>j</sub> Bild von ihm<sub>i/\*j</sub>/sich<sub>\*i/j</sub>.

Runner et al. (2003, 2006) verwenden Studien im Visual-World-Paradigma, um die Grenzen strukturgeleiteter Ansätze der Anaphernbindung und die Rolle nicht-struktureller Faktoren zu untersuchen und die Präferenzen für die Bindung von Reflexivpronomen und Personalpronomen näher zu bestimmen. Gegenstand der Untersuchungen sind PNP mit Possessor. In den Experimenten mussten die VTn mit einer Puppe (Handpuppe mit identifizierbarem Gesicht korrespondierend zu einem der möglichen Antezedenzen) entsprechend einer verbalen Anleitung ein Gesicht (von mehreren auf einem Bildschirm abgebildeten Gesichtern) berühren, das mit der anaphorischen Referenz korrespondiert. Dabei wurden die Fixationen auf den abgebildeten Gesichtern sowie die tatsächliche Entscheidung aufgezeichnet. Die Verwendung des Visual-World-Paradigmas folgt in diesem Zusammenhang der Logik, dass die Fixationen zeitlich eng mit dem gegenwärtig in Betracht gezogenen Referenten gekoppelt sind, und folglich einen Hinweis darauf geben, welcher Referent zu welchem Zeitpunkt als mögliches Antezedens berücksichtigt wird. Diese Methode wurde in in einem ersten Experiment in ihrer Validität getestet, wofür Sätze verwendet wurden, für die anhand struktureller Bindungsprinzipien eindeutige Vorhersagen getroffen werden können. Für diese Sätze wurden frühere Befunde aus Studien mit Akzeptabilitätsbeurteilung repliziert (Runner et al., 2006).

Für die Experimentalbedingungen im zweiten Experiment wurden PNP mit Possessor (*Harry*) hinsichtlich der Verwendung von Reflexiv- (Bsp. 19a und 19c) oder Personalpronomen (Bsp. 19b und 19d) variiert. Um den Einfluss der instruktiven Einleitungssätze zu kontrollieren, wurden auch diese in ihrer Reihenfolge variiert. Gemäß der Bindungstheorie sollte in der PNP mit Possessor eine Bindungsdomäne herrschen, in der der Possessor als Antezedens (*Harry*) für das Reflexivpronomen (*himself*) zugänglich ist, nicht aber für das Personalpronomen (*him*), für das entweder das Satzsubjekt (*Joe*) oder die im Einleitungssatz genannte Person (*Ken*) als Antezedens erwartet wird.

- (19) a. Pick up Joe. Look at Ken. (*Einleitende Instruktion*)  
 Have Joe touch Harry's picture of him.

---

Ansatz (Pollard & Sag, 1992; Reinhart & Reuland, 1993) entwickelt, in dem Reflexivpronomen in dieser Struktur als freie nicht-reflexive Elemente verwendet und nach diskurs-kontextuellen und pragmatischen Einflüssen aufgelöst werden (vgl. Diskussion in Runner et al., 2006, S. 195-200). Weitere Befunde und Erklärungsansätze zu Verletzungen bindungstheoretischer Annahmen in PNP liefern zum Beispiel Keller und Asudeh (2001) sowie Asudeh und Keller (2001), die mit Magnitude Estimation-Studien die Rolle pragmatischer Faktoren für die Auflösung von Reflexivpronomen hervorheben. Asudeh und Keller (2001) stellen dem strukturbasierten Ansatz auf dieser Grundlage einen prädikationsbasierten Ansatz gegenüber. In diesem werden Konstruktionen wie zum Beispiel *tell a story* als komplexe Prädikate aufgefasst, die die Anapher als Co-Argument zum Satzsubjekt aufnehmen. Unter der Voraussetzung, dass der Possessor nicht als Kopf der PNP aufgefasst wird, ist er in diesem Ansatz nicht mehr zwingendes Antezedens des Reflexivpronomens. Als verallgemeinertes Prinzip hierzu formulieren die Autoren das Pronoun Distribution Principle.

<sup>22</sup>Einen allgemeinen, diskurstheoretischen Ansatz, der über das Phänomen der PNP hinausragt, schlägt (MacWhinney, 2008) mit seiner Hypothese der perspektivengeleiteten Anaphernaufflösung vor.

- b. Pick up Joe. Look at Ken. (*Einleitende Instruktion*)  
Have Joe touch Harry's picture of himself.
- c. Look at Ken. Pick up Joe. (*Einleitende Instruktion*)  
Have Joe touch Harry's picture of him.
- d. Look at Ken. Pick up Joe. (*Einleitende Instruktion*)  
Have Joe touch Harry's picture of himself.

Die Ergebnisse für die Auswahl des Koreferenten – d.h. auf welches Gesicht tatsächlich gezeigt wurde – weisen auf eine schwache Befolgung struktureller Beschränkungen und damit auf Kompatibilität mit bindungstheoretischen Annahmen hin. Allerdings wurde die Kompatibilitätsannahme sowohl für Personalpronomen (etwa 11% der Fälle) als auch für Reflexivpronomen (etwa 31% der Fälle) recht häufig verletzt. Zudem findet sich hierin eine Interaktion von bindungstheoretischer Kompatibilität und Anaphertyp, die auf eine wesentlich stärkere Verletzung entsprechender struktureller Beschränkungen durch Reflexivpronomen hinweist – gegenüber einer weitgehend strukturell bedingten Auflösung im Falle der Personalpronomen gemäß Prinzip B. Dieses Ergebnis spricht für die Hypothese, dass auch PNP mit Possessor besser als Logophore erklärt werden können.

Die Ergebnisse aus den Verteilungen der aufgezeichneten Fixationen ermöglichen einen präziseren Einblick in die zeitliche Auflösung der Verarbeitung. Hier zeigt sich, dass bereits auf der frühesten Position auch der inkompatible Referent fixiert wird, während der dritte Referent nicht fixiert wird. Dies war auch dann der Fall, wenn nur die Trials berücksichtigt wurden, bei denen anschließend eine zur Bindungstheorie kompatible Auswahl erfolgte. Zudem traf dies sowohl für Reflexivpronomen als auch für Personalpronomen zu, obwohl sich letztere im Offline-Resultat eher nach dem strukturellen Prinzip beschreiben lassen. Diese Ergebnisse sprechen gegen Ansätze, die ein Primat entsprechender struktureller Bindungs-Constraints in der frühen Analyse annehmen (u.a. Nicol & Swinney, 1989; Sturt, 2003). Allerdings kann daraus keine generelle Aussage für die Verarbeitungsorganisation von Anaphern außerhalb des Phänomens der PNP abgeleitet werden.

**Formenspezifische multiple Constraints** Schließlich wäre noch die Hypothese formenspezifischer multipler Constraints zu erwähnen (form-specific multiple-constraints approach, Brown-Schmidt, Byron & Tanenhaus, 2005; Kaiser et al., 2009). Diese folgt der Annahme, dass abhängig vom Anaphertyp unterschiedliche Constraints in unterschiedlicher Gewichtung interagieren. Folglich werden unterschiedliche Effekte und unterschiedliche Details bei der Performanz und der zeitlichen Auflösung für Personalpronomen gegenüber Reflexivpronomen erwartet. Dies wurde aber bislang nur für die satzübergreifende Anaphernauflösung beobachtet. Kaiser et al. (2009) überprüfen die Hypothese formenspezifischer multipler Constraints innerhalb von Sätzen. Dazu verwenden sie zwei Offline-Studien<sup>23</sup> und dem Ansatz von Runner et al. (2006) folgend ein Eyetracking-Experiment im Visual-World-Paradigma. Gegenstand sind PNP mit und ohne Possessor. Im Design der Studien werden in der Anaphern-Bedingung der Anaphertyp (Reflexivpronomen vs. Personalpronomen) und in einer semantischen Bedingung durch das Verb die thematische

---

<sup>23</sup>Unter anderem wurde ein Experiment mit Bild-Verifizierungs-Aufgabe (Sentence-Picture-Matching-Task) durchgeführt. Bei diesem müssen die VTn beurteilen, ob ein akustisch präsentierter Satz mit einer Abbildung auf dem Bildschirm korrespondiert.

Rolle von Subjekt und Objekt (Quelle vs. Empfänger) verändert (vgl. Bsp. 20). Aufgrund von Bindungspräferenzen, die für andere Strukturen beobachtet wurden, wird eine Präferenz der Anbindung des Reflexivpronomens an das Quellen-Objekt und eine Präferenz der Anbindung des Personalpronomens an das Empfänger-Objekt erwartet. Bei einem Primat struktureller Constraints sollten diese semantischen Bedingungen aber keinen Einfluss auf die Verarbeitung nehmen.

- (20) a. Peter told John about the picture of him on the wall.  
*(Personalpronomen, Subjekt: Quelle)*
- b. Peter told John about the picture of himself on the wall.  
*(Reflexivpronomen, Subjekt: Quelle)*
- c. Peter heard from John about the picture of him on the wall.  
*(Personalpronomen, Subjekt: Empfänger)*
- d. Peter heard from John about the picture of himself on the wall.  
*(Reflexivpronomen, Subjekt: Empfänger)*

Neben der Verletzung der bindungstheoretischen strukturellen Constraints finden Kaiser et al. (2009) in den verschiedenen Studien Hinweise darauf, dass verschiedenen Anapherentypen tatsächlich eine spezifische Sensitivität für verschiedene (strukturelle oder semantische) Constraints aufweisen. Speziell Reflexivpronomen zeigen eine wesentlich höhere Sensitivität für strukturelle Constraints als Personalpronomen, die sich stärker durch die semantischen Bedingungen beeinflussen lassen. Diese Interaktion von Anapherentyp und strukturellen vs. semantischen Beschränkungen unterstützt die Hypothese formenspezifischer multipler Constraints auch innerhalb von Sätzen. Die Befunde aus dem Eyetracking-Experiment zeigen zudem, dass sich der Einfluss der semantischen Constraints nicht als später Effekt auswirkt, wie dies von Theorien einer gestuften Verarbeitung mit strukturell determiniertem Filter angenommen werden würde. Vielmehr setzt er bereits sehr früh – ab 200-250 ms nach On-Set der Anapher – als initialer Constraint der Anaphernaufflösung ein.



## 6. AnaSRN: SRNs und Referenzialität

Wir haben in Kapitel 3 die Architektur rekurrenter Netzwerke vorgestellt und ihre Verwendung in Bezug auf die Modellierung von Sprachverarbeitungsprozessen eingeführt. Diese Klasse von Netzwerken weist spezifische Eigenschaften bei der Verarbeitung sprachlicher Strukturen auf, die in Beziehung zu psycholinguistischen Daten der Satzverarbeitung gestellt werden können und möglicherweise Erklärungsansätze für die Eigenschaften menschlicher Sprachverarbeitungsperformanz bieten. Im folgenden Kapitel werden wir die Erweiterung der Architektur zur Erfassung referenzieller Aspekte in der Sequenzverarbeitung besprechen: Die Verarbeitung anaphorischer Elemente mit der Zuweisung des Antezedens. Wir werden dabei ein auf SRNs basierendes Modell entwickeln und untersuchen, das neben der Wortvorhersage auch referenzielle Zuweisung leisten soll. Diese Erweiterung trägt der Tatsache Rechnung, dass wir in SRNs mit klassischer Wortvorhersage viele Aspekte außer Acht lassen müssen, da sie jenseits der Oberflächeninformation der vorhergesagten Wortformen zu suchen sind. Im speziellen Fall der Anapher beispielsweise scheint es recht einfach festzustellen zu sein, dass nach einer syntaktischen Wortfolge auf Objektposition außer einem expliziten Nomen auch ein anaphorischer Ausdruck folgen kann. An entsprechender Stelle werden dann grammatikalisch kongruente anaphorische Formen aktiviert. Genügt dies, um die Komplexität der Verarbeitung an entsprechender Stelle vorherzusagen und eventuelle Schwierigkeiten der Integration des Wortes in die bestehende Struktur zu erklären? Wir bezweifeln dies und nehmen statt dessen an, dass zur Beurteilung und Vorhersage der Komplexität der Anapherauflösung in der Online-Satzverarbeitung auch die Auflösung der Anapher bzw. die Präzision der Bindung an einen Co-Referenten berücksichtigt werden muss.

Wir präsentieren die Ergebnisse des Projekts *AnaSRN* (Kurzform für Anapherauflösung in Simple Recurrent Networks). Ausgehend von einem früheren Modellierungsansatz mit kaskadierend kombinierten SRNs erläutern wir unseren eigenen Ansatz für die Modellierung der Anapherauflösung (vgl. Weldle, Baumann et al., 2009; Weldle, Konieczny, Müller, Wolfer & Baumann, 2009b). Wir stellen mehrere Simulationen vor, die in diesem Ansatz durchgeführt wurden. Die Ergebnisse der Simulationen erweitern frühere Befunde zu Merkmalen der Satzverarbeitung in SRNs, können in Beziehung zu verschiedenen empirischen Befunden gesetzt werden und führen zu Vorhersagen für die Online-Verarbeitung von Anaphern bei der natürlichen Sprachverarbeitung.

## 6.1. Konnektionistische Modellierung der Anaphernaufflösung

Es gibt unseres Wissens nur wenige konnektionistische Modellierungsstudien, die die Auflösung anaphorischer Referenzen direkt thematisiert haben. Ausgangspunkt unseres Projekts ist eine Modellierungsstudie mit SRNs, anhand deren das kritische Ergebnis formuliert wurde, SRNs seien nicht in der Lage, irrelevante nicht-strukturelle Faktoren und Hinweisreize zu ignorieren. Die Studie von R. Frank et al. (2005) setzt sich zum Ziel, die Eigenschaften und Grenzen von SRNs als Modell menschlicher Satzverarbeitung zu untersuchen. Als Gegenstand wählen die Autoren die Verarbeitung anaphorischer Referenzen. Diesem schreiben sie zu, dass er sowohl lexikalisch als auch strukturell komplex ist, und damit eine echte Herausforderung für die Modellklasse der SRNs darstellt.<sup>1</sup> Vor allem aus der Kritik an dieser Studie werden wir einige Voraussetzungen formulieren, die ein entsprechendes Modell der Anaphernaufflösung erfüllen sollte. Die wenigen weiteren Modelle, die die Auflösung anaphorischer Referenzen simulieren, tragen unseres Erachtens wenig oder überhaupt nichts zum von uns gewählten Ansatz und den von uns untersuchten Fragestellungen bei. Deshalb werden sie hier nur kurz skizziert.

In einem sehr frühen Ansatz stellt Reilly (1984) ein Modell der Auflösung einfacher *it*-Anaphern vor; die Merkmale des Modells sind allerdings den technischen Beschränkungen der Zeit geschuldet sehr rudimentär. In einer Untersuchung des Cascade-Correlation-Algorithmus als generativem Lernalgorithmus für konnektionistische Netzwerke<sup>2</sup> entwickeln Shultz, Buckingham und Oshima-Takane (1994) ein Netzwerk, das den Erwerb einfacher Personalpronomen der ersten und zweiten Person Singular (*ich* und *du*) simuliert. Das Netzwerk lernt auf Grundlage von Informationen über den aktuellen Sprecher, den Adressaten und den Referenten die korrekte Form vorherzusagen. Modelliert werden sollten dabei nicht Mechanismen der Anaphernaufflösung, sondern Aspekte des frühkindlichen Erwerbs und Erwerbsverlaufs einfacher pronominaler Formen. Im Rahmen des DSS-Modells (S. L. Frank et al., 2003b) wurde auch eine Erweiterung für die Simulation der Auflösung von Personalpronomen vorgestellt (vgl. Kap. 4.4). Dabei steht die Auflösung ambiger Personalpronomen aufgrund der Informationsquellen *Foregrounding* (Zugänglichkeit der Antezedenzen im aktuellen Diskurs-Modell) und *Kontext* (Aktivierung des zur Auflösung benötigten Weltwissens durch den semantischen Kontext) im Vordergrund. Anaphernaufflösung wird dabei als Spezialfall allgemeiner wissensbasierter Inferenz erklärt. Der Ansatz ist auf Personalpronomen der dritten Person Singular beschränkt und speziell auf die Auflösung ambiger Anaphern über semantische bzw. pragmatische Präferenzen orientiert. Er trifft keine Aussagen zu den strukturell orientierten Fragestellungen unseres Projekts.

---

<sup>1</sup>Die Studie von R. Frank et al. (2005) wurde nicht publiziert, ist aber öffentlich über die Homepages der Autoren zugänglich (z.B.: <http://www.cog.jhu.edu/faculty/rfrank/>, 20.11.2009), und wurde von ihnen in eigenen Arbeiten und Konferenzbeiträgen sowie von zwei anderen Arbeitsgruppen zitiert. Sie darf also als ernstzunehmender Ansatz verstanden werden. Da sie trotz einiger Kritik unsererseits als initiale Inspirationsquelle für das Projekt AnaSRN diente, ist es unabdingbar, sie hier näher zu erläutern. In weiteren Studien mit sogenannten Transformations-Netzwerken verfolgen die Autoren ähnliche Ziele der Aufdeckung von Schwächen der Architektur SRN (vgl. R. Frank & Mathis, 2007).

<sup>2</sup>Unter diesem Algorithmus entwickeln die Netzwerke während des Lernens ihre eigene Architektur.

Etwas interessanter ist aus der Perspektive unseres Ansatzes die Studie von Joanisse und Seidenberg (2003). In einer Untersuchung zu spezifischen Sprachentwicklungsstörungen (Specific Language Impairment, SLI) stellen sie ein konnektionistisches Modell vor, das die Auflösung von Personal- und Reflexivpronomen anhand grammatischer bzw. struktureller Constraints und Störungen der Verarbeitung grammatischer Strukturen simulieren soll. Sie schlagen dabei eine Alternative zur Annahme vor, SLI sei eine grammatik-spezifische Schädigung: Stattdessen wird die Störung als Folgeerscheinung gestörter Sprachwahrnehmung konzipiert, d.h. als phonologischer Effekt. Die Modellierung soll demonstrieren, dass verzerrter phonologischer Input zu einer Einschränkung der Performanz der Anaphernaufflösung bei gleichzeitiger Aufrechterhaltung der allgemeinen syntaktischen Performanz führen kann. Die Autoren beziehen ihren Ansatz auch auf MacDonald und Christiansen (2002) und schlagen als Erklärung für SLI einen negativen Einfluss der verzerrten phonologischen Form auf das Arbeitsgedächtnis bzw. auf die Aufrechterhaltung syntaktischer Konstituenten vor. Für die Simulationen verwenden Joanisse und Seidenberg (2003) eine rekurrente Architektur mit sogenannten Clean-up-Units<sup>3</sup> und eine Wortassoziations-Aufgabe, bei der Sätze als sequenziell dargebotene Wörter in phonologischer Feature-Präsentation auf die aktuelle Bedeutung des Wortes in Form einer semantischen Feature-Präsentation abgebildet werden. Während das Lexikon recht komplex gewählt wird, bleibt die syntaktische Komplexität überschaubar.<sup>4</sup>

Das Netzwerk wurde mit ambigen Personalpronomen und Reflexivpronomen getestet, deren mögliche Antezedenzen jeweils den gleichen Genus vorweisen. Zusätzlich enthielt das Test-Set nicht-ambige Sätze, in denen die beiden Anaphertypen über den Genus aufgelöst werden können. Das Modell zeigt gute Performanz für die Anaphernaufflösung, wobei die Verarbeitung von Personalpronomen besser bewältigt wird als die Verarbeitung von Reflexivpronomen. Bei verzerrtem phonologischem Input nimmt die Performanz für beide Anaphertypen deutlich ab, dabei sind die Reflexivpronomen stärker betroffen. Kontinuierlich perfekt bleiben hingegen die Testsätze mit Anaphern, die über den Genus aufgelöst werden können – ein Hinweis auf eine spezifische Störung der syntaktischen Fähigkeiten ohne Beeinträchtigung der allgemeinen Fähigkeiten. Aufgrund der Orientierung des Modells auf die Simulation von spezifischen Sprachentwicklungsstörungen werden kaum Aussagen darüber getroffen, welche Mechanismen und Einflussfaktoren die Anaphernaufflösung bedingen. Folglich bleibt der Zusammenhang zu unserem Ansatz gering.

### 6.1.1. Verarbeitung struktureller Bindungs-Constraints durch SRNs

Inwiefern ist die Verarbeitung von Anaphern lexikalisch und strukturell komplex? Auf der lexikalischen Ebene bezieht sich diese Aussage auf die variable Referenzialität anaphorischer Wortformen: Anaphern beziehen sich nicht auf einen bestimmten Referenten (wie z.B. der Eigename *Daniel*) oder auf eine bestimmte Klasse von Referenten (wie

---

<sup>3</sup>Der Hidden-Layer wird durch einen etwas kleineren vollverknüpften Kontext-Layer mit einem Gedächtnis ausgestattet.

<sup>4</sup>Größere Komplexität ist nur im lokalen Rahmen mit Ditransitiven, Präpositionalphrasen und einfachen subordinierten Nebensätzen (z.B. *Bill thinks Stan likes him*) gegeben. Ernsthaftige Fernabhängigkeiten oder Relativsätze werden nicht miteinbezogen.

z.B. *Doktorand*), sondern co-referieren nur mit einem anderen Wort auf einen Referenten. Dabei kann sich ihre Interpretation auf jedes beliebige Antezedens mit übereinstimmendem Genus und Numerus beziehen. Entsprechend erwerben wir nicht die Bedingungen der Zuweisung der Anapher zu einem bestimmten Referenten, sondern die Bedingungen zur Beschränkung der im Kontext möglichen Referenten. Auf der strukturellen Ebene ist die Aussage im Sinne strukturunabhängiger Fernabhängigkeiten begründet: Die Interpretation der Anapher kann auf ein Antezedens verweisen, das über mehrere Einheiten hinweg entfernt liegt. Dabei müssen unter Umständen konkurrierende Antezedenzen, d.h. semantisch plausible Agenten mit ebenfalls übereinstimmendem Genus und Numerus, ignoriert und für die Interpretation irrelevante syntaktische Strukturen, wie zum Beispiel eingebettete Relativsätze, überbrückt werden. Die Anapher kann also in verschiedenen strukturellen Kontexten erscheinen, muss aber unabhängig von diesen interpretiert werden.<sup>5</sup>

Wie bereits erwähnt, wurde der im folgenden vorgestellte Ansatz von R. Frank et al. (2005) als kritische Überprüfung der Leistungsfähigkeit von SRNs zur Modellierung sprachlicher Aufgaben konzipiert. Die Autoren beziehen sich dabei auf zwei Desiderate der Erforschung von SRNs als Satzverarbeitungsmodell: Zum einen decken die bislang untersuchten linguistischen Strukturen nur einen kleinen Teil der in natürlichen Sprachen beobachteten Phänomene struktureller Generalisierung ab. Zum anderen wurde bislang wenig Gewicht darauf gelegt, welche internen Strukturen und Generalisierungen in SRNs zur Gewährleistung korrekter Input-Transformationen entstehen – welche Lösung die Netzwerke also für die Verarbeitung entwickeln, und inwiefern diese Lösung den Generalisierungen beim menschlichen Spracherwerb entspricht.

Als Gegenstand wählen die Autoren die Interpretation von Reflexiv- und Personalpronomen. Diese erfordern die Berücksichtigung struktureller Constraints und Kongruenz zwischen anaphorischem Ausdruck und möglichem Antezedens. Dabei müssen mitunter wie bei der Subjekt-Prädikat-Kongruenz Fernabhängigkeiten aufgelöst werden. Allerdings folgen die Generalisierungen für Anaphern anderen Regelmäßigkeiten. Wie wir im vorausgegangenen Kapitel 5.3.1 erläutert haben, werden in strukturellen Ansätzen der Anapherauflösung für Reflexiv- und Personalpronomen komplementäre Bindungsdomänen angenommen, da die beiden Anapherntypen unterschiedlichen Constraints unterliegen (Prinzip A vs. Prinzip B, vgl. Chomsky, 1981; Fanselow & Felix, 1987). Die Bindung der Anapher an ein Antezedens wird also im Sinne einer strukturellen Auflösung von Fernabhängigkeiten verstanden. Genau genommen wollen die Autoren den Filter-Mechanismus nach global wirkenden strukturellen Constraints simulieren, der sich in der Komplementarität der Bindungsdomänen auswirkt und mögliche Antezedenzen nach grammatischen Kriterien aus diesen Bindungsdomänen auswählt.

---

<sup>5</sup>Dies trifft natürlich auch auf sehr viele andere Wortklassen in ähnlicher Weise zu. So können kategoriale Nomen wie zum Beispiel *Doktorand* ebenfalls anaphorisch verwendet werden und sich auf eine zuvor genannte konkrete Person beziehen. Strukturunabhängige Fernabhängigkeiten müssen auch für die Kongruenz zwischen Satz-Subjekt und finitem Verb beachtet werden – ganz im Gegensatz zur nicht nachvollziehbaren Behauptung von R. Frank et al. (2005, S. 6), Subjekt-Prädikat-Kongruenz sei strukturell ein lokales Phänomen. Im Unterschied zu kategorialen Nomen entsprechen Anaphern viel eher Variablen, da sie nur minimale grammatikalische Voraussetzungen an den zu bindenden Referenten stellen, und die semantischen Bedingungen für plausible Kandidaten durch den Kontext bestimmt werden. Im Unterschied zur Kongruenz von Verben kann die Auswahl möglicher Kandidaten auch über Satzgrenzen hinweg bestimmt werden.

Da das Modell die Referenzzuweisung in die normale Satzverarbeitung (d.h. die klassische Wortvorhersage-Aufgabe) eingliedern soll, genügt ein Standard-SRN nicht mehr für die Simulation, denn die Verarbeitung erfordert die Bewältigung zweier verschiedener Aufgaben und damit zweier verschiedener Teaching-Signale: Wortvorhersage und Referenzzuweisung. R. Frank et al. (2005) schlagen hierfür eine Kombination von zwei SRNs vor, die in zwei getrennten Schritten – aber aufeinander aufbauend – trainiert werden. SRN 1 erfüllt die klassische Wortvorhersage-Aufgabe. Das Trainings-Korpus wurde auf Grundlage einer probabilistischen kontextfreien Grammatik (PCFG) erstellt. Die künstliche Sprache orientiert sich an der englischen Grammatik. Das Lexikon enthält Verben, männliche und weibliche Eigennamen, unbelebte Objekte sowie Relativ-, Reflexiv- und Personalpronomen. Die Grammatik enthält neben einfachen Sätzen Objekt- und Subjekt-Relativsätze. Subjekt-Prädikat-Kongruenz wurde über eine Genus-Markierung des Verbs hergestellt, d.h. das finite Verb muss mit der Genus-Markierung des Subjekts übereinstimmen.<sup>6</sup> Nach vollständigem Training von SRN 1 wurden die Gewichte des Netzwerks eingefroren. Anschließend wurde SRN 2 auf die Referenzzuweisungs-Aufgabe trainiert. Dazu wurden die beiden Netzwerke gekoppelt, indem der rekurrente Hidden-Layer von SRNs 1 mit dem primären Hidden-Layer von SRN 2 (Kompressions-Layer vor dem rekurrenten Hidden-Layer von SRN 2) vollständig verknüpft wurde. Input für SRN 2 ist also die Aktivierung des Hidden-Layers des fixierten SRNs für den sequenziellen Wort-Input. Target für SRN 2 ist der Referent des aktuellen Worts – die Referenzzuweisung erfolgt nicht über eine Vorhersage-Aufgabe, sondern als rein assoziative Aufgabe. Der Target-Vektor des Referenz-Outputs gibt immer genau einen korrespondierenden Referenten vor, wenn als Input ein Nomen oder ein Pronomen gegeben ist. Alle anderen Wortklassen werden mit einem Non-Reference-Knoten assoziiert.

Das Fixieren der Gewichte nach dem ersten Trainingsschritt (SRN 1) wird folgendermaßen begründet: „We assume, following previous work, that this first training phase on the word prediction task will suffice to allow the network to acquire a hierarchical representation of the sentences in the training corpus.“ (R. Frank et al., 2005, S. 14). Damit wird SRN 2 dazu gezwungen, auf Grundlage der aus der Wortvorhersage extrahierten Struktur eine geeignete Struktur zur Lösung der Referenzzuweisung abzuleiten. Die von den Autoren gegebene Begründung der Trennung der Trainingsschritte und des Einfrierens der Gewichte wirkt nicht recht einleuchtend. Dies wird in den folgenden Abschnitten noch zu diskutieren sein.

### 6.1.1.1. Ergebnis: Beschränkte Generalisierungsfähigkeit

Sowohl für die allgemeine Performanz der Syntax-Komponente (Qualität der Wortvorhersage) als auch für die Referenz-Komponente stellen R. Frank et al. (2005) eine sehr gute Leistung bzw. einen sehr geringen Fehler fest. Als Qualitätsmaß der Wortvorhersage verwenden die Autoren die Kullback-Leibler-Divergenz (siehe Gleichung 3.4 in Kap. 3.2.1.1). Für das Test-Set erhalten sie damit einen durchschnittlichen Wert von  $D_{(\text{target} | \text{out})} = 0.029$ .

---

<sup>6</sup>Dadurch wird eine unnötige Erhöhung der Komplexität vermieden, die sich bei einer Umsetzung der Kongruenz über den Numerus ergeben würde (für nähere Erläuterungen siehe unten).

Der Anteil falscher Subjekt-Prädikat-Kongruenz beträgt nur 0.1% aller Fälle.<sup>7</sup> Für die Referenz der anaphorischen Ausdrücke wurde eine Genauigkeit von 97.3% ermittelt.<sup>8</sup> In der quantitative Analyse werden demzufolge also gute Ergebnisse erzielt.

In der Detailanalyse der tatsächlichen Output-Aktivierungen stellen die Autoren aber eine Reihe problematischer qualitativer Fehler fest, die im Falle komplexer Sätze bei der Anaphernauflösung nach eingebetteten Relativsätzen auftreten. Hauptkritikpunkt ist die Abhängigkeit der tatsächlichen Referenten-Aktivierung von irrelevanten linearen strukturellen Merkmalen. Beispiel 21 gibt die Ergebnisse für die Auflösung der Personal- und Reflexivpronomen in Abhängigkeit ihres syntaktischen Kontexts und den in diesem Kontext intervenierenden Agenten wieder. Während für das Personalpronomen nach einem zentral eingebetteten Objekt-Relativsatz (Bsp. 21a) nur korrekte Referenten aktiviert werden, wird nach einem zentral eingebetteten Subjekt-Relativsatz (Bsp. 21b) auf Kosten der Aktivierung des intervenierenden Relativsatz-Objekts zu gleichen Teilen auch das Subjekt des Matrix-Satzes aktiviert. Dieses sollte aber außerhalb der Bindungsdomäne des Personalpronomens liegen, die folglich verletzt wird. Bei einer Einbettungstiefe von zwei zentral eingebetteten Relativsätzen wird dieser Effekt noch verstärkt (Bsp. 21c). Diese Fehlaktivierung tritt aber nicht auf, wenn das intervenierende Relativsatz-Objekt selbst ein Personalpronomen ist (Bsp. 21d). Im Falle der Reflexivpronomen tritt ein ähnlicher Effekt ab einer Einbettungstiefe von zwei verschachtelten Relativsätzen auf: Während in Satz 21e nur das Subjekt des Matrix-Satzes (*Nate*) als Antezedens des Reflexivpronomens in Frage kommen dürfte, wird hier das Objekt des letzten Relativsatzes (*John*) ebenfalls aktiviert, obwohl er außerhalb der Bindungsdomäne des Reflexivpronomens stehen sollte.

- (21)
- a. Nate who Harold likes admires him  
*him*:  $p(\text{Harold}) = 0.59$ ,  $p(\text{John}) = 0.38$
  - b. Nate who likes Harold admires him  
*him*:  $p(\text{Harold}) = 0.30$ ,  $p(\text{John}) = 0.35$ ,  $p(\text{Nate}) = 0.33$
  - c. Nate who sees John who kisses Mary admires him  
*him*:  $p(\text{Harold}) = 0.28$ ,  $p(\text{John}) = 0.33$ ,  $p(\text{Nate}) = 0.38$
  - d. Nate who sees John who kisses her admires him  
*her*:  $p(\text{Alice}) = 0.38$ ,  $p(\text{Mary}) = 0.26$ ,  $p(\text{Sue}) = 0.34$   
*him*:  $p(\text{Harold}) = 0.53$ ,  $p(\text{John}) = 0.43$
  - e. Nate who sees Harold who kisses John admires himself  
*himself*:  $p(\text{John}) = 0.29$ ,  $p(\text{Nate}) = 0.70$
  - f. Nate admires himself

Als Grund für dieses Verhaltensmuster vermuten die Autoren eine Sensitivität für den lokalen Kontext: Im Falle einer kohärenten Subsequenz – in Satz 21b die Sequenz *Harold admires him*, in Satz 21e die Sequenz *John admires himself*, jeweils in Hauptsatzstruktur

---

<sup>7</sup>Zur Ermittlung der Fehlerrate wurde eine binäre Bewertung verwendet: Für jedes Verb wurde jeweils dann ein Fehler zugewiesen, wenn ein nicht übereinstimmender Verb-Knoten höher aktiviert war als der am höchsten aktivierte übereinstimmende Verb-Knoten.

<sup>8</sup>Wie im Falle der Verb-Genauigkeit wurde die Zuweisungs-Genauigkeit über eine binäre Bewertung ermittelt: Wenn der korrekte Referent nicht der am stärksten aktivierte war, wurde dies als Fehler gewertet.

– interferiert diese mit der globalen Interpretation und ermöglicht die Aktivierung von Antezedenzen, die in der Bindungsdomäne der lokalen Sequenz zugänglich sind. Während die Autoren solche Effekte als empirisch unwahrscheinlich ablehnen und der Architektur als systematische Fehlleistung ohne sprachpsychologisches Gegenstück anlasten, wurden entsprechende Effekte lokaler syntaktischer Kohärenz (LSK) für andere Strukturen bereits mehrfach empirisch belegt (vgl. Kap. 5.2). So schließen R. Frank et al. (2005) daraus, dass SRNs stark von linearen Generalisierungen abhängen, und unfähig sind diese linearen Abhängigkeiten zu ignorieren, selbst wenn sie für eine bestimmte Aufgabe irrelevant sind. Denn die intervenierende syntaktische Struktur der Relativsätze sollte für die Auflösung der anaphorischen Referenz keine Rolle spielen: Die Generalisierungen für deren Interpretation müsste unabhängig vom linearen Kontext erfolgen und über diesen hinweg abstrahieren bzw. die für die Generalisierung irrelevanten Differenzen im Kontext ignorieren.

Diese Abhängigkeit von linearen Generalisierungen belegen die Autoren weiterhin über verschiedene Verfahren der Analyse interner Repräsentationen (vgl. Kap. 3.2.1.1). Sie vergleichen dazu die Aktivierung der Hidden-Layer für die anaphorischen Ausdrücke in verschiedenen syntaktischen Kontexten. Anhand der Analysen wird demonstriert, dass die Clusterung der internen Repräsentationen nicht in erster Linie über die referenzierten Personen erfolgt, sondern über die Satztypen (einfacher Hauptsatz, intervenierender Subjekt-Relativsatz, intervenierender Objekt-Relativsatz).<sup>9</sup> Die Ergebnisse unterstützen die aus den Performanzergebnissen geschlossene Evaluation der Leistungsfähigkeit von SRNs: Diese Modellklasse scheint nicht dazu in der Lage zu sein, abstrakte lexikalische Generalisierungen zu erfassen, und ist unfähig, irrelevante Generalisierungen über die lineare Struktur zu ignorieren.

R. Frank et al. (2005) gehen noch einen Schritt weiter und testen die Systematizität von SRNs bei der Referenzzuweisung für ausgelassene (d.h. im Trainings-Set nicht als Antezedens erscheinende) Referenten für Reflexivpronomen. Die Aufgabe ist vergleichbar mit der Leistung der Identitätsfunktion, anhand derer G. F. Marcus (1998) die Systematizität von SRNs im „*rose-is-a-rose*“-Problem bezweifelt (vgl. 3.1.4): Kann das Netzwerk im Satz 21f für das Reflexivpronomen den korrekten Referenten *Nate* zuweisen, wenn dieser im Trainings-Set zwar als Eigenname, aber niemals als Antezedens für Reflexivpronomen auftritt? Während das Netzwerk für im Trainings-Set enthaltene Antezedenzen korrekte Aktivierungen generierte, stellte sich für das vorenthaltene Antezedens heraus, dass dieses überhaupt nicht aktiviert wurde. Während sich die Voraussetzungen von der Simulationsstudie von G. F. Marcus (1998) unterscheiden, da das Training mit vorenthaltenem Referenten auf dem vollständig trainierten SRN aus Stufe 1 basiert, zeigt sich für die Performanz also dasselbe Bild: Die Identitätsfunktion konnte nicht abgebildet werden, und folglich sind die in Stufe 1 entwickelten Generalisierungen nicht so abstrakt, wie erwartet wurde.

---

<sup>9</sup>Die Ergebnisse der Clusteranalyse und der Multidimensionalen Skalierung wurden weiterhin über eine Contribution Analysis bestätigt. Für kontrastierte Satzpaare mit unterschiedlicher syntaktischer Struktur, aber identischen Ziel-Referenten, konnte jeweils wechselseitig ein Knoten identifiziert werden, der die Qualität der Performanz bei der Referenzauflösung für den einen Satztyp deutlich reduziert, während die Performanz für den jeweils anderen Satztyp unbeeinträchtigt bleibt.

### 6.1.2. Kritik an R. Frank et al. (2005)

Dieser negativen Einschätzung der Eigenschaften von SRNs sind mehrere Kritikpunkte an der von R. Frank et al. (2005) verwendeten Simulationsmethodik entgegenzuhalten. Diese betreffen (a) die architektonische Lösung und die Trainingsprozedur, (b) Details der Simulationsmethodik, (c) Details der Auswertungsmethodik, und (d) in mehrerlei Hinsicht Merkmale der verwendeten Stimuli und der diesen zugrundeliegenden Grammatik.

Ein Problem der Architektur in Verbindung mit der darauf ausgelegten Trainingsprozedur liegt in der Trennung des Trainings für Wortvorhersage und Referenzzuweisung und dem Einfrieren der Gewichte der Wortvorhersage-Komponente nach der ersten Trainingsstufe. Die Argumentation für diese unmotivierte Trennung ist zumindest fragwürdig. Sie basiert darauf, dass das Netzwerk aus dem ersten Trainingsschritt zur Abbildung von Wortfolgewahrscheinlichkeiten eine für die Folgeaufgabe hinreichende interne Repräsentation entwickelt – wobei die Folgeaufgabe vollkommen anderer Natur ist und die Autoren durchaus erkennen, dass sich intern in erster Linie eine hierarchische Repräsentation zugunsten der Erkennung syntaktischer Merkmale entwickelt. Wir vermuten, dass die Ergebnisse der Simulation darauf zurückzuführen sind, dass die aus dem fixierten SRN 1 vorgegebenen Repräsentationen inadäquat für die Lösung der Referenzzuweisungs-Aufgabe sind. Genau genommen ist es sogar merkwürdig zu erwarten, dass eine Aufgabe auf der Basis von internen Repräsentationen erfüllt wird, die überhaupt keine der benötigten strukturellen Hinweise zur Lösung der Aufgabe bietet.<sup>10</sup> Denn die Wortvorhersage-Aufgabe beruht auf der Erfassung ganz anderer Aspekte und Unterscheidungen:

- Für die Vorhersage muss nicht zwischen einzelnen Personen unterschieden werden, denn es genügt, die Personen in ihrer Funktion als Subjekt oder Objekt des Satzes und ihrer Wirkung auf das finite Verb des Satzes (d.h. hinsichtlich der Genus-Kongruenz) zu unterscheiden.<sup>11</sup>
- Es muss keine Beziehung zwischen den belebten Personen und den Personal- und Reflexivpronomen hergestellt werden, denn diese treten unabhängig von den Personen einfach im Sinne weiterer Argumente auf, die eben nur in Objekt-Position stehen können.
- Es muss keine interne Repräsentation für den aktuellen Input ermittelt werden, sondern für den als nächstes folgenden Input (hier ein Verb oder das Nomen eines Folgesatzes).

Die internen Repräsentationen sind also einer gänzlich unterschiedlichen Klassifikation geschuldet, die keinen Rückschluss auf den referenziellen Bezug von Wörtern eröffnet. Durch die Trennung des Trainings und die Fixierung des Zustands aus dem ersten Trainingsschritt wird die Generierung von sinnvollen und angemessenen Repräsentationen für

---

<sup>10</sup>Um dies auf eine etwas polemische Weise zu umschreiben: Es würde schließlich auch als nicht nachvollziehbar betrachtet werden, wenn man einem Kind das Radfahren beibringt, um es anschließend in einen Fluß zu werfen und zu erwarten, dass es aufgrund der erworbenen Fähigkeiten schwimmen kann.

<sup>11</sup>Typischerweise erwerben SRNs die Fähigkeit zur Wortvorhersage durch die Entwicklung interner Repräsentationen, die Wortklassen reflektieren. Hierzu generalisieren sie über Einzelvertreter derselben Klasse und verlieren damit Informationen über die einzelnen Vertreter. Spezifische Informationen sind im internen Zustandsraum nur noch als marginale Unterschiede geringer Distanz zum zentralen Vertreter zugegen (vgl. Kap. 3.1.3.2).

die Referenzzuweisungs-Aufgabe (durch Restrukturierung der internen Repräsentationen) unterbunden.

Möglicherweise ebenfalls problematisch ist die Verwendung sehr großer Hidden-Layer in beiden Komponenten: In der Wortvorhersage-Komponente stehen nur 28 Input-Units insgesamt 70 Units im rekurrenten Hidden-Layer gegenüber. Während dies hier noch zu vertreten wäre, werden auch in der Referenzzuweisungs-Komponente 70 Units im Hidden-Layer verwendet, die nur noch 7 Output-Units gegenüberstehen. Es fragt sich, ob bei Hidden-Layern dieser Größe überhaupt noch eine sinnvolle Generalisierung zustande kommt. Vielmehr liegt der Verdacht nahe, dass das Netzwerk übergeneralisiert und Sequenzen sozusagen auswendig lernt. Auf solche Zusammenhänge zwischen zu groß gewählten Hidden-Layern und Übergeneralisierung (sogenanntes Overfitting) wird in der einschlägigen Literatur mehrfach hingewiesen (vgl. z.B. S. L. Frank, 2006c). Die unterstellte Abhängigkeit der SRN-Performanz von linearen Generalisierungen wäre dann nur ein Artefakt schlecht gewählter Netzwerk-Parameter.

Direkt mit der getrennten Trainingsprozedur und der Fixierung der Gewichte hat auch eine Unstimmigkeit der Grammatik zu tun. Wie erläutert, gibt es für das Netzwerk keinen Grund, für die Wortvorhersage-Aufgabe zwischen den einzelnen Personen zu unterscheiden und entsprechend distinktive interne Repräsentationen zu entwickeln (die noch dazu über längere Zeit aufrecht erhalten werden sollten). Um dennoch eine Distinktion zu erzwingen, verwenden R. Frank et al. (2005) einen Trick: Sie führen in die Grammatik neben den sechs Personen auch sechs unbelebte Objekte ein (u.a. *hotdogs* und *mangos*), die an die Personen gebunden sind. Mit diesen personalisierten Objekten kann also für jede Person genau eines der Objekte als zweites Argument auf der Objekt-Position erscheinen (so sind z.B. die *hotdogs* immer nur Satz-Objekt in Matrixsätzen, in denen *Nate* Satz-Subjekt ist).<sup>12</sup> Dadurch muss das Netzwerk eine Unterscheidung der Personen aufrecht erhalten.

Mit dem semantischen Pseudo-Constraint sind wir auch bei einem Review der verwendeten Grammatik angelangt: Auffällig ist zunächst die Verwendung einer unnatürlichen Subjekt-Prädikat-Kongruenz, nach der das Verb in Übereinstimmung zum Genus des Subjekts markiert ist (also Verben in den Fällen Maskulinum und Femininum). Im Englischen und auch im Deutschen ist das Verb nach Person und Numerus des Satz-Subjekts markiert. Im Falle der hier verfolgten Simulationsziele scheint es aber durchaus vertretbar, auf eine künstliche Markierung zurückzugreifen. Denn die Einführung des Plurals würde eine massive Erhöhung der Komplexität der Grammatik und der Modellierung<sup>13</sup> mit sich führen, die für die Fragestellung der Studie nicht zielführend wäre. Zu hinterfragen ist allerdings, ob diese künstliche Kongruenz zwingend über das Genus erfolgen muss, denn damit wird durch die Nennung des finiten Verbs die Genus-Information des Subjekts aktiviert, zu dem das Verb kongruent ist. Folgt nun auf das Verb eine Anapher, so muss diese keine Fernabhängigkeit mehr überbrücken, um an die relevante Genus-Information

---

<sup>12</sup>Es ist unbegreiflich, mit welcher Rechtfertigung die Autoren diesen Trick als semantischen Constraint verkaufen wollen, der zu zusätzlicher Komplexität der Grammatik führt.

<sup>13</sup>Unter anderem müsste eine Unterscheidung zwischen Eigennamen und kategorialen Namen eingeführt werden. Eigennamen müssten des Weiteren Kategorien zugeordnet werden und als Untermenge einer Kategorie erkennbar sein können. Ganz abgesehen davon müsste auch eine Lösung dafür gefunden werden, welche Referenten bei Verwendung des Plurals aktiviert werden sollen – erfolgt in der Referenzzuweisungs-Komponente bei Nennung eines Nomens im Plural die Aktivierung mehrerer oder gar aller Zugehöriger der Klasse oder nur eines Vertreters?

zu gelangen. Darüber hinaus weist die Grammatik einige Schwächen auf: Sie enthält keine Satzzeichen und keine Artikel. Dadurch ist die Länge der Sequenzen auch bei tieferen Satzeinbettungen sehr reduziert und die Leistung der Kongruenz über Fernabhängigkeiten unproblematisch zu erwerben. Durch die Verwendung von Eigennamen wird zudem kein Unterschied zwischen Objekten im Nominativ und im Akussativ markiert.

Zuletzt wäre noch die Verwendung der Kullback-Leibler-Divergenz als globales Fehlermaß für die Wortvorhersage-Performanz sowie die Verwendung der beschriebenen binären Bewertung zur Beurteilung des Subjekt-Prädikat-Kongruenz zu bemängeln. Wie in Kapitel 3.2.1.1 erläutert, gehen in die Kullback-Leibler-Divergenz nur die Aktivierungen ein, für die auch durch den Vergleichswert eine Aktivierung vorgesehen ist. Dadurch wird die große Menge der Informationen aus falsch aktivierten Units nicht berücksichtigt. Gleichermäßen sagt die Verwendung der binären Bewertung für die Subjekt-Prädikat-Kongruenz nicht viel über die tatsächliche Performanz aus, da er keine Informationen über die tatsächliche Verteilung der Aktivierung enthält.<sup>14</sup> Folglich ist das positive Ergebnis für die Wortvorhersage, auf dessen Grundlage die Ausgereiftheit und Eignung des ersten SRNs für die Folgeaufgabe beurteilt wurde, zu hinterfragen. Weshalb R. Frank et al. (2005) das Ergebnis nicht wie auch bei der Referenzzuweisungs-Komponente als oberflächliches quantitatives Ergebnis aufgefasst und die Aktivierungen bei der Wortvorhersage im Detail analysiert haben, bleibt ihr Geheimnis.

## 6.2. Anaphernauflösung in SRNs

Unseres Erachtens lässt sich für den Ansatz von R. Frank et al. (2005) nicht einschätzen, ob das Netzwerk bereits eine hinreichend gute Performanz entwickelt hat, da die verwendeten Fehlermaße ungeeignet sind. Noch mehr aber ziehen wir in Zweifel, dass das Modell die Aufgabe der Referenzauflösung unter den gegebenen Voraussetzungen von Architektur, Trainingsprozedur und Stimuli überhaupt lernen kann. Hieraus eine allgemeine negative Beurteilung der Leistungsfähigkeit von SRNs zu formulieren ist allerdings nicht gerechtfertigt. Ausgehend von unserer Kritik entwickeln wir einen neuen Ansatz, der eine Neugestaltung der konnektionistischen Architektur, des Trainingsaufbaus und der Grammatik bzw. der Trainingsstimuli beinhaltet. In diesem Unterkapitel stellen wir unseren Ansatz und die Modifikationen vor und geben eine Übersicht über die durchgeführten Simulationen.

### 6.2.1. Integratives holistisches Training

Wie umgehen wir das Problem der Referenzzuweisung auf Grundlage von Generalisierungen für die Vorhersage von Wortklassen? Die einfachste Lösung lautet: Indem wir die unterschiedlichen Aufgaben nicht getrennt, sondern parallel in einem einzigen Schritt trainieren. Wir verwenden zwei SRNs und verknüpfen diese in einer kaskadierenden Struktur, in der die Aktivierung vom Input bis zu beiden Output-Layern und der Fehler komplett von den Output-Layern zurück fließt, um alle Gewichte anzupassen. An der Oberfläche gleicht

---

<sup>14</sup>So könnten etwa vier der fünf maskulin-markierten Verben überhaupt nicht und das fünfte eben gerade etwas mehr als das am stärksten aktivierte feminin-markierte Verb aktiviert sein – dies würde dennoch eine vollwertige positive Bewertung für diesen Fall erzielen.

eine solche Doppel-SRN-Struktur der bereits von R. Frank et al. (2005) benutzten Architektur. Allerdings verwenden die Autoren für ihre Simulationen im Grunde zwei getrennte SRNs mit unabhängiger Gewichts-Entwicklung und starrem Interface. Damit wird jegliche Interaktion zwischen den beiden Komponenten – zwischen ihren Repräsentationen, ihren Ziel-Strukturen, ihren hierarchischen Abhängigkeiten und ihren unterschiedlichen Aufgaben – unterbunden. Dagegen scheint es sehr viel näher am natürlichen Spracherwerb, SRNs mit einer referenziellen Komponente auszustatten, die direkt mit der Komponente interagieren kann, die die Sequenz-Lern-Aufgabe erfüllt.

Die Interaktion der Komponenten wird vor allem dann wichtig, wenn das zu modellierende sprachliche Phänomen nicht allein aus den Kollokationen sprachlicher Einheiten und unabhängig von deren Bedeutung erklärt werden kann oder soll. Wir gehen davon aus, dass die Verarbeitung anaphorischer Referenzen nicht darauf reduzierbar ist, sondern eine interaktive Indizierung erfordert. Das parallele holistische Training soll integrierte interne Repräsentationen forcieren, die optimale Bedingungen bzw. Kategorisierungen für die Lösung der unterschiedlichen Aufgaben (bezüglich ihrer sequenziell-hierarchischen Abhängigkeiten) bieten. Übertragen auf die natürliche menschliche Sprachverarbeitung entspricht dies rudimentär einem situierten oder modalitätenübergreifenden Spracherwerb. Ein Ansatz mit paralleler Verarbeitung von Sprache und Referenz berücksichtigt, dass Sprache situiert erworben und in direkte Beziehung zur umgebenden physikalischen und sozialen Umwelt gesetzt wird. Dies ist der funktionale Kern von Sprache – die Indexierung von Objekten, Zuständen, Situationen und Ereignissen über sprachliche Referenz. Die Idee, SRNs zu erweitern und sie parallel-interaktiv mit unterschiedlichen Informationen und Aufgaben zu trainieren, wurde von verschiedenen Autoren vorgeschlagen (u.a. Christiansen et al., 1998; Dienes et al., 1999; Altmann & Mirkovic, 2009). Durch ihre Fähigkeit, probabilistische Hinweisreize aus mehreren Input-Quellen zu integrieren und dabei versteckte strukturelle Informationen abzuleiten, bieten sie über die Integration verschiedener Aufgaben und Informationsquellen einen Ansatz, bestimmte kognitive Leistungsmerkmale als emergente Phänomene verschiedener Constraints zu erklären.

Die Forderung der Implementierung einer interaktiv-referenziellen Verarbeitung kann auch generell auf SRNs bezogen werden. Elman (1990) weist bereits im Artikel „Finding structure in time“ auf die Besonderheit hin, die mit der Reduzierung des Spracherwerbs auf eine Wortvorhersage-Aufgabe im klassischen SRN-Modell unterstellt wird: Er vergleicht die Netzwerke dabei mit Kindern, die ihre Erstsprache vom Radiohören lernen.<sup>15</sup> Dies ist ein allgemeines Problem von SRNs, durch das die Interpretation der hier entwickelten internen Repräsentationen eher in die Nähe kollokationsbasierter Konzeptionen von Semantik (etwa HAL oder LSA, vgl. Burgess & Lund, 1997; Landauer et al., 1998) rückt. Zwar

<sup>15</sup>Dies kann offensichtlich nicht funktionieren. Hierfür gibt es einerseits empirische Befunde (vgl. Davis, 2002; Ellis, 1998), andererseits ist dies aus logischer Sicht unzureichend. Denn zwar könnte das Kind daraus Schlüsse auf Kollokationen von Spracheinheiten ziehen, doch weder die Spracheinheiten noch ihr Verbund stünde in irgendeiner verweisenden Beziehung zu Objekten, Zuständen und Situationen in der Welt, da beim Erwerb keine Indexierung darauf stattfindet. Auf dieses logische Problem der Zirkularität eines solchen isolierten Systems verweisen in Bezug auf mentale Konzepte im Allgemeinen unter anderem Searle (1980) in einem Gedankenexperiment, das als „Chinese Room“ bekannt wurde, und in einer Fortsetzung dazu Harnad (1990) mit dem sogenannten „Chinese Dictionary Problem“ (vgl. die Diskussion hierzu in Kap. 8.1). Mitunter wurde die Leistung von SRNs in diesem Zusammenhang gar als bloße Wortklassen-Identifizierung kritisiert (Steedman, 1999).

wird dadurch ein relevanter Teil des zu erwerbenden grammatischen und semantischen Wissens erfasst, aber eben noch lange nicht alles, was für den vollständigen Spracherwerb notwendig ist (vgl. z.B. die Diskussion in Glenberg & Robertson, 2000).

### 6.2.2. Architektur mit kaskadierenden SRNs

Abbildung 6.1 zeigt die Architektur, die wir im Projekt AnaSRN verwenden. Wir gehen davon aus, dass die Aufgaben Wortvorhersage und Referenzzuweisung die Erfassung und Abbildung jeweils eigener hierarchischer Abhängigkeiten mit einer eigenständigen sequenziellen Distribution erfordern. Folglich kombinieren wir hierzu zwei SRNs.<sup>16</sup> Die Wortvorhersage-Komponente des Modells arbeitet nach dem klassischen Schema, das wir ausführlich in Kapitel 3.1.2 dargestellt haben. Sie erhält am Input-Layer Sätze beliebiger Länge. Diese werden sequenziell Wort für Wort in lokalistischer Enkodierung dargeboten. Synchron zum Input wird am Output-Layer der Komponente das jeweils nächste Wort der Satzsequenz als Teaching-Signal angelegt. Anhand des Algorithmus Backpropagation Through Time wird der Fehler über den Hidden-Layer zum (Kontext- und) Input-Layer zurückpropagiert. Die Gewichte werden entsprechend angepasst. Zusätzlich nimmt der Fehler aus der Referenzzuweisungs-Komponente Einfluss auf die Gewichtsangpassung.

Die Referenzzuweisungs-Komponente erhält ihren Input über eine vollständige Verknüpfung vom Hidden-Layer der Wortvorhersage-Komponente. Grundlage der Verarbeitung ist also ebenfalls der sequenziell-lokalistisch dargebotene Satz des Input-Layers, der allerdings über den Hidden-Layer der Wortvorhersage-Komponente vermittelt wird. Ansonsten erfolgt der Aktivierungsfluss in dieser Komponente wie in einem normalen SRN. Der eigentliche Unterschied zeigt sich am Output-Layer: Hier wird synchron zum sequenziellen Wort-Input der lokalistisch enkodierte aktuelle Referent (d.h. ein distinkter Knoten je Referent) des aktuell am Input anliegenden Wortes als Teaching-Signal angelegt.<sup>17</sup> Anhand Backpropagation Through Time wird der Fehler über den Hidden-Layer der Referenzzuweisungs-Komponente weiter zum Hidden-Layer der Wortvorhersage-Komponente bis zurück zum Input-Layer propagiert. Die Gewichte werden entsprechend angepasst.

Die Aufgaben der beiden Komponenten unterscheiden sich nicht nur hinsichtlich der zu assoziierenden Informationen zwischen Input und Target, sie sind innerhalb eines Zeitschritts im Aktivierungsfluss auch zeitlich asynchron: Die Aufgabe der Komponente für die Wortvorhersage ist es, die Wörter vorherzusagen, die nach einem Wort in Abhängigkeit seines sequenziellen Kontexts möglich sind ( $t_{\text{target}} = t + 1_{\text{input}}$ ). Die Aufgabe der Referenzzuweisungs-Komponente ist die Assoziation eines Referenten synchron zum Wort-Input ( $t_{\text{target}} = t_{\text{input}}$ ). Während es abhängig von der Natur der Aufgaben auch

---

<sup>16</sup>Wenngleich hierfür zu prüfen wäre, ob die Aufgaben nicht gleichermaßen durch eine Kombination eines SRNs mit einer einfacheren Architektur (z.B. einem FFN) gelöst werden können. In der Tat haben wir dies untersucht und kommen auf qualitativ sehr ähnliche Ergebnisse – die Leistung nach quantitativer Messung liegt nur geringfügig hinter der der Doppel-SRN zurück.

<sup>17</sup>Für direkt benannte Referenten (*Polizist*) wird der entsprechende Referent (*Polizist*) aktiviert. Im Falle von Verben, Artikeln, nicht-belebten Objekten und Konjunktionen wird ein gemeinsamer Non-Referent-Knoten aktiviert. Im Falle anaphorischer Ausdrücke erfolgt die Zuweisung folgendermaßen: Bei Reflexivpronomen (*sich*) und Relativpronomen (*der/die/den*) wird der strukturbedingt korrekte Referent (*Polizist*) aktiviert, im Falle von Personalpronomen (*ihn/sie*) wird per Zufall einer der strukturbedingt möglichen Referenten (*Bäcker*,  $p = 0.5$ /*Lehrer*,  $p = 0.5$ ) aktiviert.

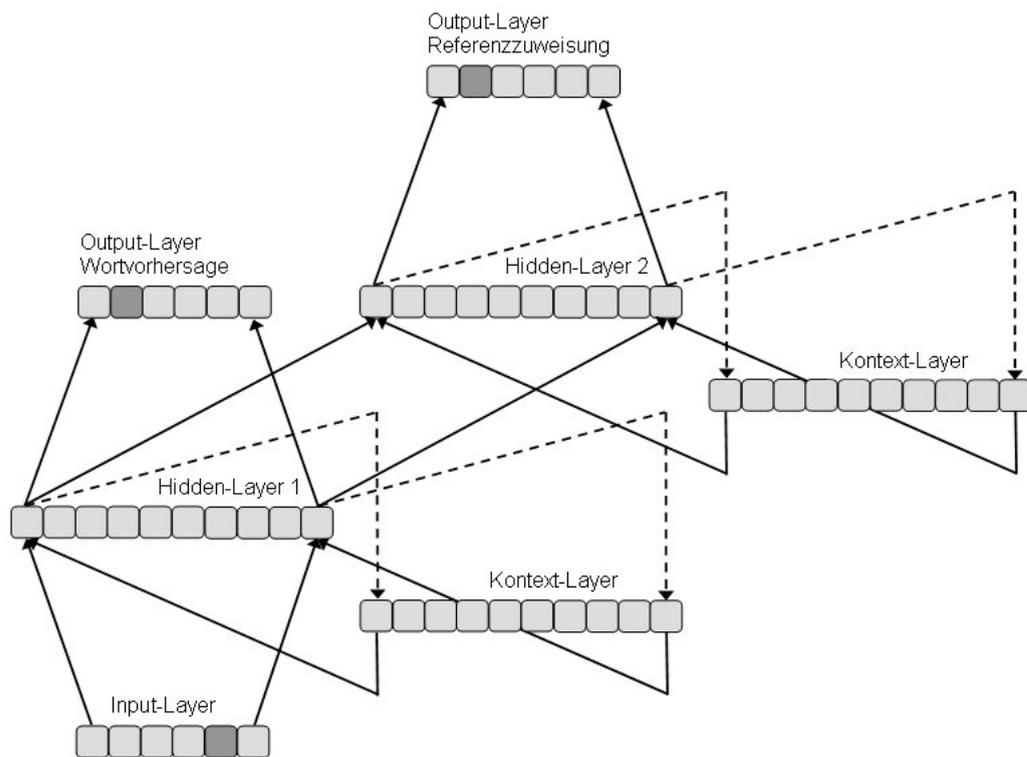


Abbildung 6.1.: Architektur mit kaskadierenden SRNs, die parallel-holistisch trainiert werden. Dabei erfolgt die Wortvorhersage in SRN-Komponente 1 und Referenzzuweisung in SRN-Komponente 2.

sinnvoll sein kann, zwei Vorhersage-Aufgaben zu integrieren (vgl. z.B. Dienes et al., 1999; Altmann, 2002; Altmann & Mirkovic, 2009), ist diese Variante mit asynchronen Aufgaben für unsere Implementierung zwingend, denn schließlich geht es um die Assoziation von Wörtern mit Referenten bzw. im Falle der Pronomen um die kontextgebundene Indizierung eines Referenten. Würden wir hier stattdessen parallel zur Wortvorhersage eine Referentenvorhersage implementieren, so müssten für alle möglichen Folgewörter (alle nominalen und pronominalen Formen) alle möglichen Referenten aktiviert werden, da die tatsächlich folgende Wortform ja noch nicht bekannt ist. Dies wäre für unsere Fragestellung und unseren Modellierungsgegenstand nicht sinnvoll. Hier ist festzuhalten, dass es sich bei unserem Ansatz um ein Modell der Referenzierung bei der Satzverarbeitung handelt und explizit nicht um ein Modell des Sprachverstehens – es wird keine komplexe situationale Repräsentation von Bedeutung aufgebaut. Folglich werden keine situationalen Zustände, keine Relationen zwischen Zuständen und keine Zustandsveränderungen abgebildet, für die die Implementierung einer Vorhersage-Aufgabe sinnvoll wäre.

### 6.2.3. AnaSRN: Simulationsstudien

Auf der Grundlage unseres neuen Modells haben wir eine Modellierungsstudie zur Anaphernaufflösung entwickelt. Das Modell dient dabei einerseits der Untersuchung von allgemeinen Fragestellungen zu konnektionistischen Ansätzen der Sprachverarbeitung (Anaphernaufflösung, Variablenbindung, Systematizität), andererseits der Modellierung und Vorhersage empirischer Daten im Sinne einer konnektionistischen Psycholinguistik (Online-Anaphernaufflösung, Auflösung von Fernabhängigkeiten, Spracherwerb). Die Eigenschaften unserer Simulationen werden neben der Gestaltung der Architektur und der Trainingsprozedur durch Variationen der Trainingsbedingungen sowie durch die Realisierung der Grammatik bedingt, anhand derer die Trainingsstimuli erzeugt werden und über die die strukturellen Eigenschaften der künstlichen Sprache bestimmt wird:

- Wir trainieren das Netzwerk parallel-interaktiv: Es bekommt von Anfang an die komplette Information und muss alle Aufgaben in jeweils einem Aktivierungsfluss lösen. Das bedeutet auch, dass alle internen Strukturierungen, die zur Lösung der unterschiedlichen Aufgaben entstehen, sich gegenseitig beeinflussen können. Ziel ist die Untersuchung von Eigenschaften und Unterschieden der Strukturierung.
- Wir haben verschiedene Grammatiken erarbeitet, die die Schwächen der Vorgängerstudie umgehen und so die Qualität der Stimuli steigern: Das Material enthält komplexere syntaktische Strukturen und erfordert die Verarbeitung ernstzunehmender Fernabhängigkeiten. Darüber hinaus wurden künstliche Eigenschaften reduziert,<sup>18</sup> die die Auflösung von Fernabhängigkeiten vereinfachen oder die Performanz auf andere Art und Weise verfälschen würden. In verschiedenen Versionen der Grammatik haben wir unter anderem die Realisierung der Subjekt-Prädikat-Kongruenz und der Satzzeichen variiert.
- Mit gezielt strukturierten Test-Sets und geeigneten Analysemethoden werden strukturelle Effekte und Interferenz-Quellen identifiziert, die auf die Performanz einwirken. Im Vordergrund stehen Effekte struktureller Komplexität und lokaler syntaktischer

---

<sup>18</sup>Dies betrifft vor allem die personalisierten Objekte und die genusbasierte Subjekt-Prädikat-Kongruenz.

Kohärenz sowie systematische Differenzen der Auflösung verschiedener Typen anaphorischer Referenz.

- Durch zusätzliche Tests werden weitere Eigenschaften der Architektur identifiziert: Mit vorenthaltenem Feedback für die Referenzzuweisungs-Komponente für einzelne Referenten soll die Systematizität bei der Extrapolation der Generalisierungen auf Gegenstände außerhalb der Trainingsstimuli geprüft werden. Simulationen gänzlich ohne Referenzzuweisung dienen dem Vergleich der entwickelten internen Strukturen mit den Strukturen der holistisch trainierten Netzwerke.

Folgende Simulationsstudien wurden durchgeführt: Studie 1 befasst sich mit einer Simulation, bei der die Grammatik sehr einfach gehalten wurde. Studie 2 zeigt eine Simulation mit einer maximal komplexen Grammatik. Studie 3 umfasst drei Simulationen, die zum einen die grammatischen Eigenschaften aus Studie 2 variieren und zum anderen das Feedback der Referenzzuweisungs-Komponente manipulieren. Mit der höheren Komplexität der Grammatiken werden die Anforderungen an das Netzwerk höher und die distributionalen Eigenschaften realistischer und damit brauchbarer für Vergleiche mit empirischen Daten. Die Architektur von AnaSRN und Teile der mit ihr durchgeführten Modellierungen wurden von Weldle, Baumann et al. (2009) und Weldle, Konieczny et al. (2009b) vorgestellt.

### 6.2.3.1. Grammatiken und Lexika

Für die Simulationsstudien haben wir zwei Basis-Grammatiken und daraus abgeleitete Varianten erstellt, wobei die zweite Grammatik mit Blick auf die grammatischen Anforderungen an das Netzwerk komplexer und besser kontrolliert ist als die erste. So erfordert sie unter anderem die Verarbeitung ernstzunehmender Fernabhängigkeiten, bietet ein breiteres Spektrum verschiedener syntaktischer Strukturen und reduziert künstlich vereinfachende Hinweisreize. Die erste Grammatik hingegen hält sich weitgehend an die Struktur der Vorgängerstudie von R. Frank et al. (2005) und modifiziert diese in Orientierung an grammatischen Merkmalen des Deutschen. Die verschiedenen Realisierungen der Grammatik sollen Aufschluss über mögliche Wechselwirkungen zwischen strukturellen Merkmalen der Stimuli und Effekten in der Performanz für verschiedene Satztypen geben.<sup>19</sup>

Die beiden PCFGs sind an den Merkmalen der deutschen Grammatik orientiert, weichen jedoch auch in mehreren Punkten deutlich von dieser ab. In den Tabellen 6.1 und 6.2 werden die Strukturen der PCFGs in reduzierter Form dargestellt.<sup>20</sup> In Tabelle 6.3 werden die Konstituenten der Grammatik aufgeführt. In den abschließenden Tabellen 6.4 und 6.5 werden mit den Terminalen die Lexika zu den Grammatiken dargestellt.

<sup>19</sup>Wie im Vorwort bereits angekündigt, wurde auf die Einbindung eines Anhangs verzichtet. Die elektronische Version des Anhangs im PDF-Format kann auf Anfrage per E-Mail an [helmut.weldle@googlemail.com](mailto:helmut.weldle@googlemail.com) zugesandt. In diesem Anhang findet sich auch ein Abschnitt mit den vollständigen Versionen der hier vorgestellten SLG-Grammatiken.

<sup>20</sup>Um die Grammatik einigermaßen überschaubar und nachvollziehbar zu halten, verwenden wir einige vereinfachende Darstellungen. Vor allem werden Hilfskonstruktionen ausgelassen, mit denen etwa die Markierung der Interpunktion durch Punkt oder Komma kontrolliert wird. Zudem verzichten wir auf die Aufführung der Constraints für die Auswahl der Terminalen. Die Darstellung der Wahrscheinlichkeiten für die Auswahl verschiedener Phrasenstrukturregeln behalten wir jedoch bei, da sie Aufschluss über die distributionalen Eigenschaften der Korpora bieten.

<b>Produktion</b>	→	<b>Phrasenstrukturregel</b>
S	→	HS . ( $p = 0.7$ )   topNS . ( $p = 0.3$ )
HS	→	NPnom VP ( $p = 0.6$ )   NPnom VP NS ( $p = 0.15$ )   NPnom VP KonP HS ( $p = 0.15$ )   NPnom VP KonP topNS ( $p = 0.1$ )
NS	→	KonH NPnom NPakk V ( $p = 0.7$ )   KonH NPnom NPakk V KonP HS ( $p = 0.2$ )   KonH NPnom NPakk V KonP topNS ( $p = 0.1$ )
topNS	→	KonH NPnom NPakk V nachHS
nachHS	→	V NPnom NPakk ( $p = 0.7$ )   V NPnom NPakk KonP HS ( $p = 0.2$ )   V NPnom NPakk KonP topNS ( $p = 0.1$ )
RC	→	PNrel-O NPnom V ( $p = 0.5$ )   PNrel-S NPakk V ( $p = 0.5$ )
NPnom	→	Nnom ( $p = 0.8$ )   Nnom RC ( $p = 0.2$ )
NPakk	→	Nakk ( $p = 0.4$ )   Nobj ( $p = 0.1$ )   PNpers ( $p = 0.15$ )   PNrefl ( $p = 0.15$ )   Nakk RC ( $p = 0.2$ )
VP	→	V NPakk

Tabelle 6.1.: Grundzüge der PCFG der Simulationsstudie 1 in vereinfachter Form, ohne Constraints.

<b>Produktion</b>	→	<b>Phrasenstrukturregel</b>
S	→	HS . ( $p = 0.75$ )   topNS . ( $p = 0.25$ )
HS	→	NPnom VP ( $p = 0.6$ )   NPnom VP, NS ( $p = 0.05$ )   NPnom VP, KonP HS ( $p = 0.3$ )   NPnom VP, KonP topNS ( $p = 0.05$ )
NS	→	KonH NPnom NPakk Adv V ( $p = 0.8$ )   KonH NPnom NPakk Adv V, KonP HS ( $p = 0.1$ )   KonH NPnom NPakk Adv V, KonP topNS ( $p = 0.1$ )
topNS	→	KonH NPnom NPakk Adv V, nachHS
nachHS	→	V NPnom Adv NPakk ( $p = 0.7$ )   V NPnom Adv NPakk, KonP HS ( $p = 0.2$ )   V NPnom Adv NPakk, NS ( $p = 0.05$ )   V NPnom Adv NPakk, KonP topNS ( $p = 0.05$ )
RC	→	PNrel-O NPnom Adv V ( $p = 0.5$ )   PNrel-S NPakk Adv V ( $p = 0.5$ )
NPnom	→	Det Nnom ( $p = 0.7$ )   Det Nnom, RC, ( $p = 0.3$ )
NPakk	→	Det Nakk ( $p = 0.4$ )   Det Nobj ( $p = 0.1$ )   PNpers ( $p = 0.15$ )   PNrefl ( $p = 0.15$ )   Det Nakk, RC, ( $p = 0.2$ )
VP	→	V Adv NPakk

Tabelle 6.2.: Grundzüge der PCFG der Simulationsstudie 2 in vereinfachter Form, ohne Constraints.

Symbol	Konstituente
S	Satz
HS	Hauptsatz
NS	Nebensatz
topNS	vorangestellter Nebensatz
nachHS	nachgestellter Hauptsatz
RC	Relativsatz
NPnom	Nominalphrase, nominativ
NPakk	Nominalphrase, akkusativ
VP	Verbalphrase
Nnom	Nomen, nominativ
Nakk	Nomen, akkusativ
Nobj	Nomen, akkusativ, unbelebt
V	Verb, finit
Det	Artikel, bestimmt
Adv	Adverb
PNpers	Personalpronomen
PNrefl	Reflexivpronomen
PNrel-O	Relativpronomen, Objekt
PNrel-S	Relativpronomen, Subjekt
KonH	Konjunktion, hypotaktisch
KonP	Konjunktion, parataktisch

Tabelle 6.3.: Erläuterung zu den Phrasenknoten der beiden PCFGs der Simulationsstudie.

Zentrale Unterschiede zwischen den beiden Grammatiken und Lexika betreffen vor allem (a) die Verwendung einer unterschiedlichen Form von Subjekt-Prädikat-Kongruenz – genus-abhängig gegenüber willkürlich zugeordnet und damit genus-unabhängig, (b) die Auslassung gegenüber der Verwendung von Artikeln anhand derer Nominativ und Akussativ unterschieden werden können, und (c) die Auslassung gegenüber der Verwendung von Adverbien.

Bevor wir die Eigenschaften der mit diesen Grammatiken erzeugten Korpora besprechen, erläutern wir noch einige nebensächliche Entscheidungen: (a) Wir haben in unseren Grammatiken die Verwendung von unbelebten Objekten als mögliche Argumente im Akkusativ beibehalten. Im Unterschied zur Vorgängerstudie wurden diese Objekte allerdings nicht personalisiert, d.h. jedes belebte Objekt kann Agent einer Handlung mit jedem der unbelebten Objekte sein. Die Verwendung von personalisierten Objekten als Trick um die Unterscheidbarkeit der Referenten aufrechtzuerhalten, ist durch die direkte Interaktion der Architekturkomponenten nicht mehr nötig. (b) Es mag zunächst verwundern, dass auch die Akussativ-Objekte mit Kongruenzmarkern versehen wurden. Diese dienen in erster Linie der Bestimmung der Kongruenz für das Verb in rechtsverzweigenden Subjekt-Relativsätze. (c) Die Wahrscheinlichkeiten der Folgekonstituenten und die relativen Wahrscheinlichkeiten der Satztypen basieren nicht auf Korpuswahrscheinlichkeiten. Sie beruhen auf intuitiven Einschätzungen der relativen Häufigkeit verschiedener Haupt- und Nebensatztypen und Satzanbindungen. Zudem wurden sie darauf ausgelegt, Verarbeitungsunterschiede für verschiedene Satzstrukturen nicht durch unterschiedliche Frequenzen (etwa für Subjekt- und Objekt-Relativsätze) zu manipulieren.

Konstituenten	Kongruenz	Terminalen
Nomen, nominativ	mask	<i>daniel   lars   sascha</i>
	fem	<i>bettina   wibke   sarah</i>
Nomen, akkusativ	mask	<i>daniel   lars   sascha</i>
	fem	<i>bettina   wibke   sarah</i>
Nomen, akkusativ, unbelebt		<i>computer   kaffebecher   zigarette   tasche</i>
Verb, finit	mask	<i>kratzt<sub>mask</sub>   sieht<sub>mask</sub>   zeichnet<sub>mask</sub>   amüsiert<sub>mask</sub>   langweilt<sub>mask</sub></i>
	fem	<i>kratzt<sub>fem</sub>   sieht<sub>fem</sub>   zeichnet<sub>fem</sub>   amüsiert<sub>fem</sub>   langweilt<sub>fem</sub></i>
Personalpronomen		<i>ihn   sie</i>
Reflexivpronomen		<i>sich</i>
Relativpronomen, Objekt		<i>den.rpo   die.rpo</i>
Relativpronomen, Subjekt		<i>der.rps   die.rps</i>
Konjunktion, hypotaktisch		<i>als   während</i>
Konjunktion, parataktisch		<i>und</i>

Tabelle 6.4.: Lexikon zur PCFG mit genus-abhängiger Subjekt-Prädikat-Kongruenz.

Konstituenten	Kongruenz	Terminalen
Nomen, nominativ	agr1	<i>baecker1   handwerker1   soldatin1   koechin1</i>
	agr2	<i>forscher2   minister2   aerztin2   richterin2</i>
Nomen, akkusativ	agr1	<i>baecker.akk1   handwerker.akk1   soldatin.akk1   koechin.akk1</i>
	agr2	<i>forscher.akk2   minister.akk2   aerztin.akk2   richterin.akk2</i>
Nomen, akkusativ, unbelebt		<i>computer   kaffebecher   zigarette   tasche</i>
Verb, finit	agr1	<i>kratzt<sub>agr1</sub>   sieht<sub>agr1</sub>   zeichnet<sub>agr1</sub>   langweilt<sub>agr1</sub></i>
	agr2	<i>kratzt<sub>agr2</sub>   sieht<sub>agr2</sub>   zeichnet<sub>agr2</sub>   langweilt<sub>agr2</sub></i>
Artikel, bestimmt		<i>der   die   den   die.akk</i>
Adverb		<i>gerade   jetzt</i>
Personalpronomen		<i>ihn   sie</i>
Reflexivpronomen		<i>sich</i>
Relativpronomen, Objekt		<i>den.rpo   die.rpo</i>
Relativpronomen, Subjekt		<i>der.rps   die.rps</i>
Konjunktion, hypotaktisch		<i>als   während</i>
Konjunktion, parataktisch		<i>und</i>

Tabelle 6.5.: Lexikon zur PCFG mit willkürlicher Subjekt-Prädikat-Kongruenz.

Außer den hier aufgeführten Grammatiken und Lexika haben wir weitere Versionen erstellt. Zentrale Unterscheidungsmerkmale der verschiedenen Versionen sind (a) die Gestaltung der Subjekt-Prädikat-Kongruenz, (b) die Verwendung oder Nicht-Verwendung unbelebter Objekte, (c) die Anzahl der unterscheidbaren Referenten, zum Beispiel eine Erhöhung auf 16 Referenten in einer erweiterten Version, (d) die Verwendung oder Nicht-Verwendung von Kommata als Interpunktionszeichen. Die meisten dieser Modifikationen erbrachten im Vergleich zu den beiden Basis-Grammatiken keine Performanz-Unterschiede. Deshalb werden wir auf nähere Erläuterungen der Stimuli und der damit erzielten Ergebnisse verzichten.

### 6.2.3.2. Trainingskorpora

Mit den PCFGs werden folgende Satztypen generiert: Einfach-transitive Hauptsätze und temporale Nebensätze in beliebig großen hypotaktischen und parataktischen Satzreihungen. Temporale Nebensätze können dabei auch vorangestellt auftreten (mit entsprechender Topikalisierung des finiten Verbs im folgenden Hauptsatz). An alle definiten Nominalphrasen, d.h. alle Nomen im Nominativ oder Akkusativ innerhalb aller Satzarten (außer Pronominalformen), können des Weiteren in beliebiger Einbettungstiefe Subjekt- und Objekt-Relativsätze angebunden werden. Die Grammatik weist mit Parataxen, Hypotaxen und Voranstellungen eine hohe Komplexität auf und erfüllt das Kriterium der Produktivität bzw. der Rekursivität. Die zweite Grammatik verfügt zudem über eine komplexere Morphosyntax und ein umfangreicheres Lexikon: Sie enthält unter anderem Artikel (mit Genus- und Kasus-Markierung), Konjunktionen (in Abhängigkeit der Satzreihungsform) und Adverbien. Der Kasus der Nomen ist markiert und die Auflösung der Reflexivpronomen ist komplexer, da diese im Unterschied zum Englischen keine Genus-Markierung tragen.

Wir unterscheiden verschiedene Korpora, die aus den Grammatiken generiert werden. Unser Hauptaugenmerk liegt auf der Untersuchung der Studien mit den Korpora 1 und 2. Die Korpora 3a, 3b und 3c dienen der Untersuchung spezieller Fragestellungen, die einzelne Aspekte der Stimuli sowie generelle Funktionsweisen der Architektur betreffen.

**Korpus 1: Einfach mit Genus-Kongruenz** Korpus 1 wird über Grammatik 1 erzeugt. Die Subjekt-Prädikat-Kongruenz ist über die Genus-Markierung realisiert (in Entsprechung zu R. Frank et al., 2005). Wir unterscheiden die Kongruenz-Marker Maskulinum (*\_mask*) und Femininum (*\_fem*): Verben können also in den beiden Konjugationen *mask* und *fem* gebildet werden, die mit dem Genus des Satz-Subjekts übereinstimmen müssen.

**Korpus 2: Komplex mit künstlicher Kongruenz** Korpus 2 wird über Grammatik 2 erzeugt. Die Subjekt-Prädikat-Kongruenz wird als künstliche Kongruenz realisiert, die unabhängig von der Genus-Markierung der Nomen ist. Wir unterscheiden die Kongruenz-Marker 1 (*\_agr1*) und 2 (*\_agr2*). Je der Hälfte der maskulinen und der femininen Nomen wird Kongruenz 1 zugeordnet, der anderen Hälfte entsprechend Kongruenz 2. Die Verben können entsprechend in den beiden Konjugationen *agr1* und *agr2* gebildet werden. Im Unterschied zu Korpus 1 ist Korpus 2 um zwei weitere Merkmale erweitert: Es verfügt über Artikel, die in Genus und Kasus mit ihrem Objekt übereinstimmen müssen. Es verfügt zudem über Kommata, über die die Interpunktion für Haupt- und Nebensatz-Grenzen markiert wird.

**Korpus 3a: Reduziert mit künstlicher Kongruenz** Korpus 3a entspricht Korpus 2, enthält jedoch keine Kommata und damit weniger explizite Hinweise auf die syntaktische Struktur.

**Korpus 3b: Partiiell reduzierte Version** Korpus 3b entspricht ebenfalls Korpus 2, jedoch wurde über die Constraints der Grammatik die Referenzzuweisung so verändert, dass ausgewählte Referenten nicht als Antezedens von Reflexivpronomen erscheinen.

**Korpus 3c: Partiiell reduzierte Version** Korpus 3c ist identisch zu Korpus 2. Der Unterschied besteht in dieser Simulation in der Veränderung der Architektur, die ohne Referenzzuweisung trainiert wird.

Die von uns gewählte Form der Subjekt-Prädikat-Kongruenz ermöglicht es, die im Deutschen gegebene Markierung des Verbs nach Person und Numerus des Satz-Subjekts zu umgehen. Wie wir in Kapitel 6.1.2 erläutert haben, vermeiden wir damit die Einführung einer großen Anzahl von Details, die für unsere Fragestellung irrelevant sind, und verhindern damit eine unnötige und nicht zielführende Erhöhung der Komplexität sowohl für die Grammatik als auch für die Referenzzuweisungs-Komponente. Wir hatten allerdings die Genus-Kongruenz insofern problematisiert, als dass damit auf der Position des finiten Verbs die Genus-Information des Subjekts aktiviert wird. In der Folge könnte die Auflösung der nachgestellten Anaphern beeinflusst werden. Die Lösung der Subjekt-Prädikat-Kongruenz in Korpus 2 mit einer genus-unabhängigen Kongruenz-Markierung ist zwar ebenfalls künstlich, sie verhindert aber eine Konfundierung der Auflösung der Subjekt-Prädikat-Kongruenz mit der Genus-Markierung der Nomen. Dadurch, dass sie unabhängig von den Zuständen der Genus-Markierung erfolgt, entspricht sie zudem eher der tatsächlich gegebenen Kongruenz über die Markierung von Person und Numerus. Deshalb erscheint uns diese Lösung sehr viel näher an natürlicher Sprache als die Lösung über eine Genus-Kongruenz. Die zusätzliche Erweiterung der Grammatik durch satzinterne Interpunktionszeichen ermöglicht die Untersuchung von Effekten durch zusätzliche Informationen der Satzstruktur, die beispielsweise durch Satzzeichen oder bei gesprochener Sprache durch prosodische Merkmale gegeben sind.

Unter Verwendung des Programms SLG (Simple Language Generator, Rohde, 1999b) wurden auf Grundlage der PCFG Trainingskorpora mit einem Umfang von 20.000 Sätzen generiert und für die Simulation aufbereitet.<sup>21</sup> Die unterscheidbaren Wortformen werden dabei über lokalistische Vektoren dargestellt, d.h. jeweils eine Unit repräsentiert ein Wort. Die Interpunktions-Marker (EOS für Punkte, COMMA für Kommata) werden in der Simulation

---

<sup>21</sup>Die Aufbereitung des Korpus zur Ableitung der Referenzzuweisung erfolgt über mehrere Schritte: (a) Erzeugung des Korpus anhand der PCFG mit SLG. Dabei sind die terminalen Knoten auch hinsichtlich ihres zugehörigen Referenten markiert, d.h. alle nominalen und pronominalen Formen enthalten einen Verweis auf den Referenten der korrespondierend aktiviert werden soll. (b) Trennung von Wortformen und Referenzmarkierungen für die Erzeugung der tatsächlichen Trainingskorpora. Erstellung der Input-Vektoren aus dem Wort-Korpus sowie durch Versetzung um eine Stelle der Target-Vektoren der Wortvorhersage-Komponente. Erzeugung der Target-Vektoren für die Referenzzuweisungs-Komponente aus dem Referenten-Korpus und Zuordnung zum korrespondierenden Wort-Input. (c) Berechnung der Wortfolge-Wahrscheinlichkeiten für die Test-Korpora für die Auswertungen auf Grundlage der PCFG mit SLG. (d) Da die Wahrscheinlichkeiten für die zusammengesetzten Formen mit Wort und korrespondierender Referenz angegeben werden, berechnen wir über verschiedene Transformierungen der SLG-Dateien und der Output-Dateien die korrespondierenden Wahrscheinlichkeiten für einzelne Wörter und die Zuweisungs-Wahrscheinlichkeiten für die Referenzzuweisung.

wie Wörter behandelt und werden dementsprechend durch eigene Vektoren repräsentiert. Die Referenten-Repräsentationen für die Referenzzuweisungs-Komponente sind ebenfalls lokalistisch enkodiert. Hinzu kommt hier die Repräsentation eines Non-Reference-Markers. Dieser wird immer dann aktiviert, wenn ein Wort keine Referenzzuweisung ermöglicht – also korrespondierend zu Verben, Artikeln, Satzzeichen, Adverbien und Konjunktionen.

### 6.3. Simulationsstudie 1

In Simulation 1 testen wir anhand des vergleichsweise einfach gehaltenen Korpus 1, inwiefern die Architektur in verschiedenen Konstruktionen Anaphern verarbeiten und das korrekte Antezedens erfassen kann. Darüber hinaus untersuchen wir die Performanz der Wortvorhersage-Komponente, um etwa Komplexitätsunterschiede zwischen Subjekt- und Objekt-Relativsätzen (vgl. z.B. MacDonald & Christiansen, 2002) zu überprüfen. Vor allem aber geht es darum festzustellen, welche der Befunde der Vorgängerstudie von R. Frank et al. (2005) sich in unserem weiterentwickelten Modell bestätigen lassen und welche Aussagen über die Performanz-Eigenschaften von SRNs zurückgewiesen werden können. Für unsere Analysen verwenden wir geeignete alternative Fehlermaße für die Wortvorhersage und die Referenzzuweisung. Effekte durch verschiedene syntaktische Strukturen oder LSK werden wir näher auf ihre Herkunft untersuchen. Dabei wird zum einen die Systematik der Fehler anhand von Detail-Analysen von Interferenz-Quellen identifiziert (vergleichbar dem Vorgehen von Konieczny, Ruh & Müller, 2009). Zum anderen werden die internen Repräsentationen der beiden Hidden-Layer untersucht, um die Generalisierungen zu überprüfen, auf denen die Lösungen der verschiedenen Aufgaben beruhen, sowie um irrelevante lineare Generalisierungen auszuschließen.

#### 6.3.1. Trainings- und Test-Stimuli

Das generierte Trainings-Korpus umfasst 20.000 Sätze mit insgesamt 211.559 Wörtern. Daraus ergibt sich eine durchschnittliche Satzlänge von 10,6 Wörtern. Einschließlich Satzzeichen enthält der längste Satz 63 Wörter, der kürzeste vier Wörter. Zur Untersuchung der Performanz der Netzwerke wurde manuell ein Test-Korpus nach den grammatischen Vorgaben der PCFG erstellt. Die hierin verwendeten Satzstrukturen werden in Beispiel 22 aufgeführt.<sup>22</sup> Für jeden der abgebildeten Satztypen wurden Varianten mit verschiedener Auswahl und Reihenfolge der Nomen und Verben erstellt<sup>23</sup>, insgesamt enthält das Test-Korpus 168 Sätze.

- (22) a. Daniel<sup>mask</sup><sub>i</sub> kratzt<sup>mask</sup> sich<sub>i</sub>/ihn<sup>\*</sup><sub>i</sub>/sie<sup>\*</sup><sub>i</sub>/Wibke<sup>fem</sup>.  
 b. Als Daniel<sup>mask</sup><sub>i</sub> den<sub>i</sub> Wibke<sup>fem</sup><sub>j</sub> sieht<sup>fem</sup> sich<sub>i</sub>/<sup>\*</sup><sub>j</sub>/ihn<sup>\*</sup><sub>i</sub>/<sub>j</sub>/sie<sup>\*</sup><sub>i</sub>/<sub>j</sub> kratzt<sup>mask</sup> langweilt<sup>fem</sup> Sarah<sup>fem</sup> Sascha<sup>mask</sup>.  
 c. Als Daniel<sup>mask</sup><sub>i</sub> der<sub>i</sub> Wibke<sup>fem</sup><sub>j</sub> sieht<sup>mask</sup> sich<sub>i</sub>/<sub>j</sub>/<sup>\*</sup><sub>i</sub>/<sub>j</sub>/sie<sup>\*</sup><sub>i</sub>/<sub>j</sub> kratzt<sup>mask</sup> langweilt<sup>fem</sup> Sarah<sup>fem</sup> Sascha<sup>mask</sup>.

<sup>22</sup>Die Indizes markieren die Koreferenz von Wörtern und damit die Beziehung zwischen Anapher und Antezedens. Indizes mit Asterisk zeigen eine unzulässige Antezedens-Beziehung.

<sup>23</sup>Satztyp 1 erhält 24 Sätze, die Satztypen 2 bis 5 jeweils 36 Sätze.

- d. Daniel<sup>mask</sup><sub>i</sub> den<sub>i</sub> Wibke<sup>fem</sup><sub>j</sub> sieht<sup>fem</sup>  
 kratzt<sup>mask</sup> sich<sub>i/\*j</sub>/ihn<sub>\*i/j</sub>/sie<sub>\*i/\*j</sub>.
- e. Daniel<sup>mask</sup><sub>i</sub> der<sub>i</sub> Wibke<sup>fem</sup><sub>j</sub> sieht<sup>mask</sup>  
 kratzt<sup>mask</sup> sich<sub>i/\*j</sub>/ihn<sub>\*i/j</sub>/sie<sub>\*i/\*j</sub>.

Als ersten Satztyp verwenden wir einfache Hauptsätze mit personaler oder anaphorischer Referenz auf der Objekt-Position (Bsp. 22a). Diese stellen sozusagen die Baseline, über die wir überprüfen können, ob die grundlegenden Aufgaben überhaupt gelernt wurden. Als Kernproblem der fehlerhaften Referenz-Aktivierung in der Studie von R. Frank et al. (2005) haben wir LSK-Effekte ermittelt (vgl. Kap. 6.1.1.1). Während Strukturen mit LSK im Englischen sehr einfach konstruiert werden können, müssen wir hier für die am Deutschen orientierte Grammatik eine andere Lösung finden.<sup>24</sup> Um eine vergleichbare Struktur für die Untersuchung von Interferenz-Effekten der lokalen Struktur sicherzustellen, haben wir folgenden Satztyp verwendet: Ein vorangestellter Nebensatz mit zentral eingebettetem Objekt-Relativsatz (d.h. mit Anbindung an das Subjekt des Nebensatzes) mit anschließendem Reflexiv- oder Personalpronomen auf der Objekt-Position des Nebensatzes (vgl. Bsp. 22b). Dieser Satztyp enthält eine LSK, die mit dem Subjekt des Objekt-Relativsatzes beginnt und die Grenze des Relativsatzes überschreitend das Pronomen mit einschließt. Dies resultiert in einer lokal kohärenten Hauptsatzform, in Satz 22b beispielsweise in der Sequenz *Wibke<sup>fem</sup> sieht<sup>fem</sup> sich*. Dies entspricht ziemlich genau der Struktur, die in der Vorgängerstudie zum dargestellten Effekt der fehlerhaften Referenz-Aktivierung führte. Als Vergleichsstruktur verwenden wir in vorangestellte Nebensätze eingebettete Subjekt-Relativsätze (Bsp. 22c). In diesen ist das Objekt des Relativsatzes nicht zwingend zum Relativsatz-Verb kongruent, denn dieses richtet sich nach dem Subjekt. Damit besteht die Möglichkeit, dass die lokale Subsequenz durch die nicht passende Kongruenz-Markierung nicht vollständig lokal kongruent ist (z.B. *Wibke<sup>fem</sup> sieht<sup>mask</sup> sich*). Als weitere Satztypen wurden Hauptsätze mit zentral eingebettetem Objekt-Relativsatz (Bsp. 22d) und Subjekt-Relativsatz (Bsp. 22e) aufgenommen. In diesen ist die Kohärenz der lokalen Subsequenz vollständig unterbunden, da nach dem Relativsatz erst ein weiteres Verb (das Matrix-Verb) folgt: Die lokale Sequenz *Wibke sieht kratzt sich* entspricht keiner gültigen Hauptsatzstruktur.

Mit den aufgenommenen Satztypen vergleichen wir Subjekt- und Objekt-Relativsätze in verschiedenen Einbettungen. Insbesondere für die Versionen mit Relativsatz im Nebensatz erwarten wir LSK-Effekte auf die Anaphernauflösung, die sich folgendermaßen auswirken sollten: Im Falle des Reflexivpronomens sollte in der Subsequenz der global unzulässige im Relativsatz genannte Referent (in Satz 22b: *Wibke*) lokal als vermeintliches Satz-Subjekt interpretiert und deshalb als mögliches Antezedens berücksichtigt werden. Dies sollte vor allem auf Kosten des global korrekten Antezedens, das Subjekt des Matrixsatzes (*Daniel*), erfolgen. Im Falle des Personalpronomens sollte das global unzulässige Matrixsatz-Subjekt (*Daniel*) als Antezedens berücksichtigt werden, da in der lokalen Subsequenz kaum oder gar nicht erkannt wird, dass es diese Satzfunktion erfüllt. Zugleich erwarten wir, dass der

<sup>24</sup> Anders als im Englischen ändert sich im Deutschen die Wortstellung im Nebensatz zur Verbletzstellung des finiten Verbs. Dies gilt gleichermaßen für Subjekt- als auch für Objekt-Relativsätze: Innerhalb dieser beiden Varianten ändert sich die Reihenfolge von Subjekt und Objekt (erkennbar an der Kasus-Markierung), nicht aber die Reihenfolge der Objekte gegenüber dem finiten Verb, das stets auf finaler Position stehen muss.

im Relativsatz genannte Referent (*Wibke*) als vermeintlich neues Satz-Subjekt der lokalen Subsequenz nur noch reduziert berücksichtigt wird.

Neben der Untersuchung der Auswirkung der LSK auf die Anapherauflösung ermöglicht die Kontrastierung der Satztypen weiterhin die Untersuchung eines Phänomens, das in der psycholinguistischen Literatur ausführlich diskutiert wurde: In verschiedenen Sprachen wurde ein Komplexitätsunterschied zwischen Subjekt- und Objekt-Relativsätzen (zugunsten des ersteren) festgestellt. Solche Komplexitätsunterschiede zwischen verschiedenen Relativsatztypen sowie zwischen verschiedenen Arten der Relativsatz-Einbettung waren auch bereits Gegenstand verschiedener konnektionistischer Simulationen auf SRN-Basis (u.a. Elman, 1991; Christiansen, 1994; Christiansen & Chater, 1999b; MacDonald & Christiansen, 2002; Rohde, 2002; Konieczny & Döring, 2003; Konieczny & Ruh, 2003; Konieczny, Ruh & Müller, 2009; Wells et al., 2009, vgl. Kap. 3.2). Die Aufgaben-Integration in unserer Architektur und die hierfür vom Netzwerk auf den verschiedenen Output-Layern abgebildeten Aktivierungsmuster ermöglichen es uns, mehr Informationen über die Performanz des Netzwerks und mögliche Interaktionen zwischen Informationsstrukturen bzw. Auswirkungen wechselseitiger Constraints zu untersuchen. Damit stellt unsere Architektur eine Möglichkeit dar, das Spektrum analysierbarer sprachlicher Aspekte und die Möglichkeiten ihrer Untersuchung auszuweiten. Deshalb ist es interessant, diese Komplexitätsunterschiede aus der Perspektive unseres Modells zu analysieren.

### 6.3.2. Technische Details zur Simulation

Die verwendete Architektur umfasst wie in Kapitel 6.2.2 erläutert zwei kaskadierend verschaltete SRNs, die parallel-interaktiv trainiert werden. Außer den genannten rekurrenten Layern wurden keine zusätzlichen Kompressions-Layer verwendet. Die SRNs wurden mit dem Algorithmus Backpropagation Through Time (Werbos, 1990) trainiert. Als Batch-Größe wurde dabei der Wert 1 gewählt, d.h. nach jedem vollständigen Satz werden die Gewichte aktualisiert. Standardmäßig erhalten die Hidden-Layer eine sigmoide Aktivierungsfunktion, die Kontext-Layer sind mit einer linearen Aktivierungsfunktion versehen. Der Fehler wird anhand Kreuzentropie bemessen. Zudem erhalten die Output-Layer eine Softmax-Funktion, die die Summe des Outputs auf 1.0 normalisiert. Die Größe des Input-Layers und des Output-Layers für die Wortvorhersage entspricht der Größe des Lexikons für Grammatik 1 (einschließlich Satzzeichen), diese erhalten damit jeweils 33 Units. Die Größe des Output-Layers für die Referenzzuweisung ist abhängig von der Menge unterscheidbarer Referenten (einschließlich des Non-Reference-Markers), für Grammatik 1 also 7 Units. Die Simulationen wurden mit der Modellierungs-Software LENS (Rohde, 1999a) umgesetzt.<sup>25</sup>

<sup>25</sup>Wir haben vorab verschiedene Software-Pakete getestet, unter anderem tlearn (Elman, 1992), JavaNNS (I. Fischer, Hennecke, Bannes & Zell, 2002), Emergent (Aisa, Mingus & O'Reilly, 2008) und JOONE (Marrone, 2007). Das schließlich verwendete Open-Source-Programm LENS von Rohde (1999a) wurde aus verschiedenen Gründen gewählt: Es hat vergleichsweise kurze Berechnungszeiten (dank seiner Implementierung in C) und implementiert sehr viele Algorithmen, Aktivierungs- und Fehlerfunktionen, Parameter und Netzwerk-Typen. Dabei ist es sehr flexibel einsetzbar und durch seine Ansteuerung an der Oberfläche mit der Script-Sprache TCL sehr komfortabel. Bedauerlich ist, dass es wie die meisten anderen oben genannten Programme nicht aktualisiert und gewartet wird.

Wir haben vorab für die Simulationen für alle unterschiedlichen Korpora mit verschiedenen Größen der Hidden-Layer experimentiert. Das Verhalten des Netzwerks ist gegenüber einem breiten Spektrum verschiedener Größen der Hidden-Layer der beiden Komponenten robust. Unsere im Folgenden getroffene Auswahl reflektiert also keine besonderen Vorzüge dieser Konfiguration. Unabhängig von der unterschiedlichen Größe der Input- und Output-Layer für die verschiedenen Grammatiken wurde in allen Netzwerken für die Wortvorhersage-Komponente die Größe des rekurrenten Hidden-Layers auf 80 Units festgelegt. Der rekurrente Hidden-Layer der Referenzzuweisungs-Komponente enthält immer 40 Units. Die einheitliche Größe der internen Layer über alle Simulationen soll die Vergleichbarkeit zwischen den Simulationen herstellen. Auch für die Trainings-Parameter haben wir über alle Simulationsstudien hinweg verschiedene Konfigurationen getestet. Innerhalb des Rahmens von Werten, die aus unserer Erfahrung als sinnvoll erachtet werden, ist die Performanz der Netzwerke ebenfalls robust gegenüber Variationen. Wir haben uns für eine mittlere Parametereinstellung entschieden, die für alle im Folgenden vorgestellten Simulationsstudien des Projekts AnaSRN verwendet wurden: Für die Lernrate wurde  $\epsilon = 0.15$  gewählt, das Momentum wurde mit dem Faktor *momentum* = 0.3 festgelegt, die initiale Randomisierung der Gewichte erfolgte innerhalb der Grenze von 0 und *initialweightrange* = 0.1.

Wir haben 10 Netzwerke für jeweils 100 Epochen trainiert. Eine Epoche entspricht bei unserem Vorgehen einem Umfang von 10.000 Sätzen. Damit wird das verwendete Trainings-Korpus mit 20.000 Sätzen nach jeweils zwei Epochen einmal vollständig durchlaufen. Die Reihenfolge des Zugriffs auf die Trainingssätze erfolgt innerhalb einer Epoche mit vollständiger Permutation.<sup>26</sup> Standardmäßig werden zudem nach jedem Satz die Kontext-Layer gelöscht. Zu diesem Trainingsablauf sind zwei Dinge anzumerken: (a) Der Trainingsumfang von jeweils 100 Epochen mag sehr hoch erscheinen. In der Tat zeigt die Wortvorhersage-Komponente bereits nach 10 Epochen eine zufriedenstellende Performanz. Für die Referenzzuweisungs-Komponente mussten wir allerdings feststellen, dass der Fehler auf den Anaphernpositionen während der ersten Epochen sehr hoch blieb und sich erst in späteren Epochen stabilisierte. (b) Es mag seltsam erscheinen, dass wir nicht vollständige Trainingsdurchläufe von 20.000 Sätzen als Epochen bezeichnen. Dies erfolgt hier aus Gründen der Pragmatik: Durch die Einteilung des Korpus in zwei kleinere Epochen mit dementsprechend geringerer Beispiellanzahl werden in den frühen Epochen die schnell veränderlichen Performanzmerkmale und Veränderungen erfassbar. Zugleich bleibt der Vorteil des mit 20.000 Sätzen umfangreicheren und dadurch hinsichtlich der Satzstrukturen besser verteilten Korpus erhalten.

### 6.3.3. Ergebnisse

In den folgenden Absätzen präsentieren wir die Ergebnisse der Simulationsstudie 1. Als Kennwerte verwenden wir in der Regel das arithmetische Mittel und die Standardabweichung für den globalen Fehler der Output-Aktivierung. Für die Sätze, die im Mittelpunkt unserer Untersuchung stehen, wird der globale Fehler nach Epochen gruppiert für die Wörter des Satzes abgetragen. Für einzelne interessante Satz-Positionen werden wir die tatsächliche lokale Output-Aktivierung im Detail analysieren, um Fehlerquellen zu identifi-

---

<sup>26</sup>Die Permutation beugt der Memorierung des Korpus und der damit verbundenen Übergeneralisierung vor.

zieren. Abschließend werden wir die Struktur der Hidden-Layer der beiden Komponenten analysieren. Als globales Fehlermaß verwenden wir für die Wortvorhersage-Komponente den GPE über normalisierte Output-Aktivierungen auf Basis der absoluten Wahrscheinlichkeit aus der zugrundeliegenden PCFG. Für die Referenzzuweisungs-Komponente verwenden wir ein daraus abgeleitetes Maß – den Reference Assignment Error (RAE). Er berechnet sich identisch zum GPE (vgl. Kap. 3.2.1.1):  $RAE = 1 - \frac{hits}{hits + false\ alarms + misses}$ . Einziger Unterschied zum GPE ist die Grundlage der Differenzberechnung für hits, false alarms und misses. Diese werden nicht aus der Differenz der Aktivierung der Wörter zu ihrer Übergangswahrscheinlichkeit nach der PCFG berechnet, sondern über die Differenz der Aktivierung der Referenten zur Wahrscheinlichkeit ihrer Aktivierung nach globalen strukturellen Bindungs-Constraints ermittelt.

Für die detaillierten Analysen der Output-Aktivierung ziehen wir Positionen heran, die aus folgenden Gründen von besonderem Interesse sind: (a) Positionen, für die wir unseren Hypothesen folgend hohe Fehlerwerte und spezifische Effekte erwarten, (b) Positionen, die aufgrund von Fernabhängigkeiten der dort möglichen Wörter oder aufgrund untypischer Satzstrukturen höhere Fehlerwerte erwarten lassen, und (c) Positionen, die in den Abbildungen des globalen Fehlers (mitunter unerwartet) hohe Fehlerwerte oder eine langsame Verringerung des Fehlers aufweisen. Neben den Aktivierungen für relevante Wortklassen, Einzelwörter, Referentenklassen oder Einzelreferenten tragen wir in diesen Diagrammen zusätzlich die absolute Wahrscheinlichkeit der Wörter nach der PCFG ab. Sowohl die Wahrscheinlichkeiten als auch die Aktivierungen summieren sich jeweils zum Gesamtwert 1.0.

### 6.3.3.1. Performanz bei der Wortvorhersage

Für die Wortvorhersage-Komponente zeigt sich schon nach wenigen Epochen eine sehr gute Performanz mit niedrigem Fehler über alle Sätze und Satzpositionen. Nach etwa 40 Epochen zeigt sich keine nennenswerte Entwicklung des Fehlers mehr. Das arithmetische Mittel über das gesamte Test-Set und alle Positionen beläuft sich nach Epoche 40 auf  $\bar{x}_{GPE_n} = 0.119$  ( $\sigma = 0.053$ ), der Mittelwert des Vorhersagefehlers für alle Verben<sup>27</sup> auf  $\bar{x}_{GPE_n} = 0.119$  ( $\sigma = 0.041$ ). Folglich wurden alle Fernabhängigkeiten gut gelernt. Zur detaillierten Analyse betrachten wir die Entwicklung während der ersten 10 Epochen. Auch hier ist die Fehlerquote bereits für alle Wortpositionen (einschließlich der Verben) zufriedenstellend<sup>28</sup>, die Fehlertendenzen sind aber besser zu unterscheiden und zu interpretieren. Zudem ergibt sich damit eine bessere Vergleichbarkeit mit den Ergebnissen früherer Simulationsstudien (etwa Christiansen & Chater, 1999b; MacDonald & Christiansen, 2002; Konieczny, Ruh & Müller, 2009). In den Diagrammen für den globalen Fehler (GPE) wird der Fehler gruppiert nach Epochen für die verschiedenen Testsatz-Strukturen wie-

<sup>27</sup>Dies ist insbesondere deshalb interessant, weil für die Verben Fernabhängigkeiten zum Satz-Subjekt erworben werden müssen. Das Test-Set enthält zudem einen großen Anteil an Fernabhängigkeiten über eingebettete Relativsätze hinweg. Ein niedriger Fehler für die Verben spricht also für eine gute Generalisierung über komplexe grammatische Strukturen.

<sup>28</sup>Fehler über alle Positionen nach Epoche 10:  $\bar{x}_{GPE_n} = 0.154$  ( $\sigma = 0.081$ ). Fehler über alle Verben nach Epoche 10:  $\bar{x}_{GPE_n} = 0.135$  ( $\sigma = 0.047$ ). Interessanterweise ist der Fehler für die pronominalen Wortformen (Reflexiv-, Relativ- und Personalpronomen) nach dieser Epoche noch sehr hoch:  $\bar{x}_{GPE_n} = 0.318$  ( $\sigma = 0.088$ ).

dergegeben. Die Datenpunkte repräsentieren das arithmetische Mittel für die Performanz über alle Netzwerke und alle Sätze des Satztyps für eine bestimmte Epoche (die jeweils korrespondierende Epoche kann in den Abbildungen der jeweils nebenstehenden Legende entnommen werden), der Fehlerbalken zeigt den Standardfehler. Auf der x-Achse sind die Satzpositionen, d.h. die Wörter des jeweiligen Satzes, in der Reihenfolge ihres Erscheinens abgetragen. Die y-Achse zeigt den normalisierten globalen Fehler GPEn.

**Einfache Hauptsätze** Bereits im Fehlerverlauf für einfache Hauptsätze (ohne Abb.) findet sich ein interessantes und unerwartetes Merkmal, das sich auch bei den anderen Satztypen zeigt: Die einzige Position, die einen leicht erhöhten Fehler und einen etwas langsameren Verbesserungsverlauf zeigt, ist die der pronominalen Formen *sie*, *ihn* und *sich*. Aus der Aktivierungsverteilung lässt sich der Grund dafür aber nur erahnen: Die vergleichsweise hohe Anzahl zu aktivierender Units scheint eine leichte Verschiebung der Verteilung mit sich zu bringen. Sie führt zu einer anfangs zu starken und darauf folgend zu geringen Aktivierung der Eigennamen. Die unterschiedliche Aktivierung von eigentlich gleich wahrscheinlichen Units führt in den Einzelsätzen zu einem erhöhten GPE, der bei Betrachtung der gemittelten Aktivierungsverteilungen nicht mehr nachvollziehbar ist. Auch bei den komplexeren Satztypen finden wir einen erhöhten Fehler auf dieser Position. In den frühen Epochen beruht dieser zum Teil auf anderen Ursachen. Der in den späten Epochen bleibende Fehler muss aber ebenfalls auf die zugegebenermaßen etwas unbefriedigende Erklärung der ungleichmäßigen Verteilung der Aktivierung zurückgeführt werden.

**Komplexe Nebensätze** Die beiden folgenden Abbildungen zeigen vorangestellte Nebensätze mit eingebettetem Subjekt-Relativsatz (Abb. 6.2) und eingebettetem Objekt-Relativsatz (Abb. 6.3). Der Fehler für Epoche 10 zeigt zum einen leicht erhöhte Werte im Verlauf der Relativsätze und wie bereits bei den einfachen Hauptsätzen einen übereinstimmend hohen Fehler auf der Position der pronominalen Formen. Betrachtet man die Entwicklung und Fehlerspitzen während der früheren Epochen, so zeigen sich mehrere auffallende Positionen: Erhöhte Werte zeigen sich für das Nomen im Relativsatz sowie das Relativsatz-Verb, für das Pronomen im Matrixsatz sowie mit besonders hohem Anfangsfehler und besonders schneller Fehlerreduzierung für das Matrix-Verb, und schließlich für das Verb und das Subjekt des folgenden Hauptsatzes. Wir besprechen zuerst die hinsichtlich ihrer Verursachung weniger interessanten Fehler, und greifen anschließend die Fehler auf, die für die weitere Diskussion interessant sind und die wir deshalb mit zusätzlichen Abbildungen erläutern.

Der Fehler auf dem Subjekt bzw. Objekt des Relativsatzes (*wibke.akk*, *wibke.nom*) ist in den beiden Satztypen trotz der Ähnlichkeit der Fehlerkurven auf unterschiedliche Fehlerquellen zurückzuführen. Im Falle des Subjekt-Relativsatzes beruht er vor allem auf einer zu geringen Aktivierung der auf dieser Position möglichen Reflexiv- und Personalpronomen und auf einer falschen Aktivierung der Verben sowie geringfügiger des Relativpronomens. Im Falle des Objekt-Relativsatzes beruht er hingegen auf einer zu geringen Aktivierung des Satz-Objekts (Eigennamen), dessen Fehlaktivierung aber keiner eindeutigen Überaktivierung einer anderen Wortklasse (geringfügig für die Verben) gegenübersteht. Diese Fehler sind am ehesten der nicht-kanonischen Wortreihenfolge in Relativsätzen zuzuweisen, deren Regelmäßigkeit aber mit weiterem Training erworben wird. Der Fehler auf

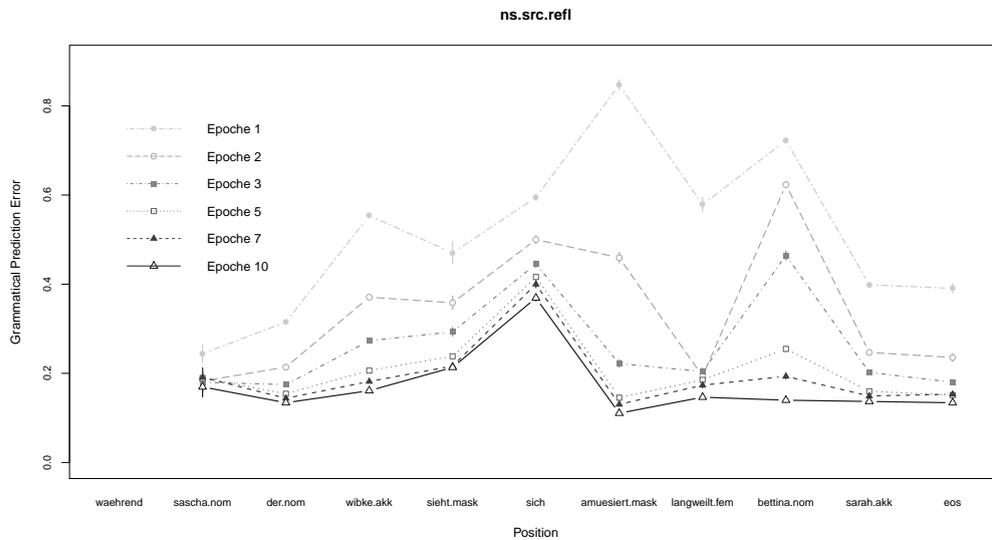


Abbildung 6.2.: Simulation 1. Verlauf und Entwicklung des globalen Fehlers der Wortvorhersage-Komponente, gruppiert nach Epoche: GPEn über Wortpositionen für vorangestellte Nebensätze mit zentral eingebettetem Subjekt-Relativsatz (Sätze mit bedingter LSK).

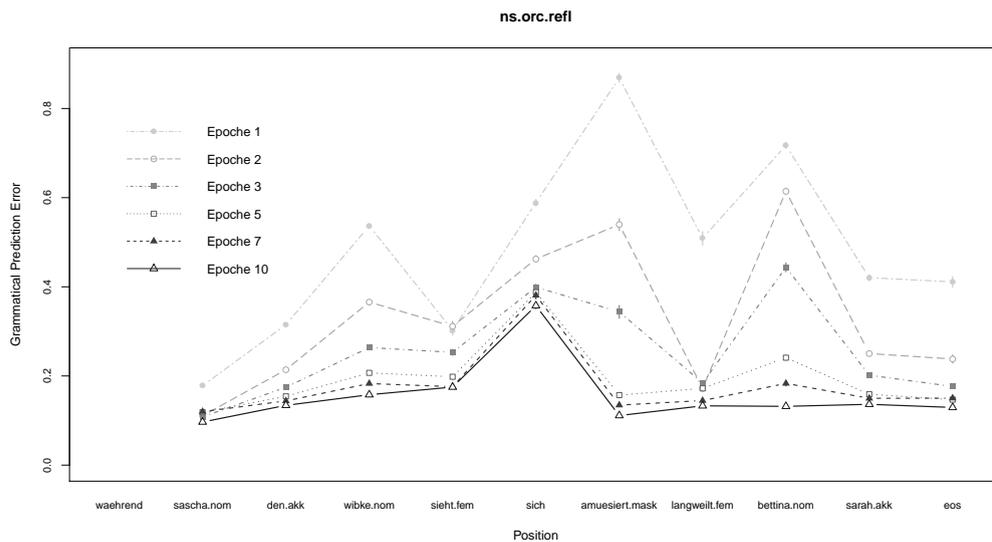


Abbildung 6.3.: Simulation 1. Verlauf und Entwicklung des globalen Fehlers der Wortvorhersage-Komponente, gruppiert nach Epoche: GPEn über Wortpositionen für vorangestellte Nebensätze mit zentral eingebettetem Objekt-Relativsatz (Sätze mit LSK).

dem Relativsatz-Verb (*sieht.mask, sieht.fem*) lässt sich für beide Satztypen der gleichen Ursache zuweisen: Zum einen einer anfangs starken Aktivierung der Verben mit falscher Kongruenz, zum anderen einer zu geringen Aktivierung der Relativpronomen, die auf dieser Position folgen könnten. Die Aktivierung der falschen Verben wirkt sich bei den Subjekt-Relativsätzen stärker aus, was an der etwas größeren Distanz zum Subjekt liegen dürfte, das die Kongruenz vorgibt. Nach kurzer Zeit zeigt sich in beiden Relativsatztypen eine zu hohe Aktivierung der korrekten Verben, die auf Kosten der Relativpronomen erfolgt – die als mögliche Fortsetzung mit einem weiteren eingebetteten Relativsatz ebenfalls korrekt wären. Dieser Fehler ist permanent und erklärt den auch nach Epoche 10 stabil erhöhten Fehlerwert. Er weist möglicherweise auf die beschränkte Fähigkeit von SRNs zur Erfassung mehrfach rekurrenter Satzstrukturen hin. Der kurzzeitig hohe Fehler auf dem Verb des folgenden Hauptsatzes (*langweilt.fem*) kommt vor allem auf Kosten der Verben durch die starke Aktivierung des EOS-Markers sowie der Konjunktionen zustande, die hier nicht möglich wären. Das Netzwerk scheint hier noch nicht erfasst zu haben, dass nach einem vorangestellten Nebensatz andere Regelmäßigkeiten gelten als nach einem gewöhnlichen Nebensatz. Bereits nach Epoche 2 wird dieser Unterschied aber erkannt. Das bedeutet, dass das Netzwerk eine Regelmäßigkeit erkennt, die auf einem Unterschied beruht, der über 8 Satzpositionen zurückliegt (dem Unterschied, ob vor dem Nebensatz ein Hauptsatz erscheint oder nicht). Der Fehler auf dem folgenden Satz-Subjekt des Hauptsatzes (*Bettina*) beruht auf der anfangs starken falschen Aktivierung der maskulinen Nomen. Das Netzwerk muss erst lernen, dass sich in der seltenen Wortreihenfolge des nachgestellten Hauptsatzes die Kongruenz der Nomen nach dem zuvor erscheinenden Verb richten muss.

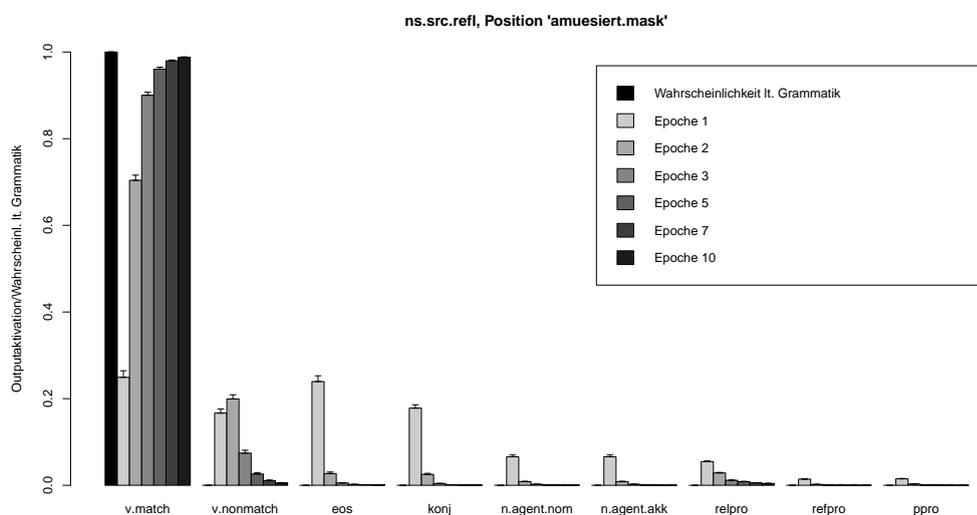


Abbildung 6.4.: Simulation 1. Details der Aktivierung im Vergleich zur absoluten Wortfolge-Wahrscheinlichkeit in ihrer Entwicklung über die Epochen für die Position des Matrix-Verbs im Nebensatz nach eingebettetem Relativsatz (*Als Daniel der Sascha sieht sich kratzt ...*).

Besonders interessant sind für uns die Positionen des Pronomens im Matrixsatz (im Beispielsatz *sich*) sowie des Matrix-Verbs (*amuesiert.mask*). Der frühe hohe Fehler auf der Position des Matrix-Verbs (hier stellvertretend dargestellt für den Subjekt-Relativsatz, Abb.

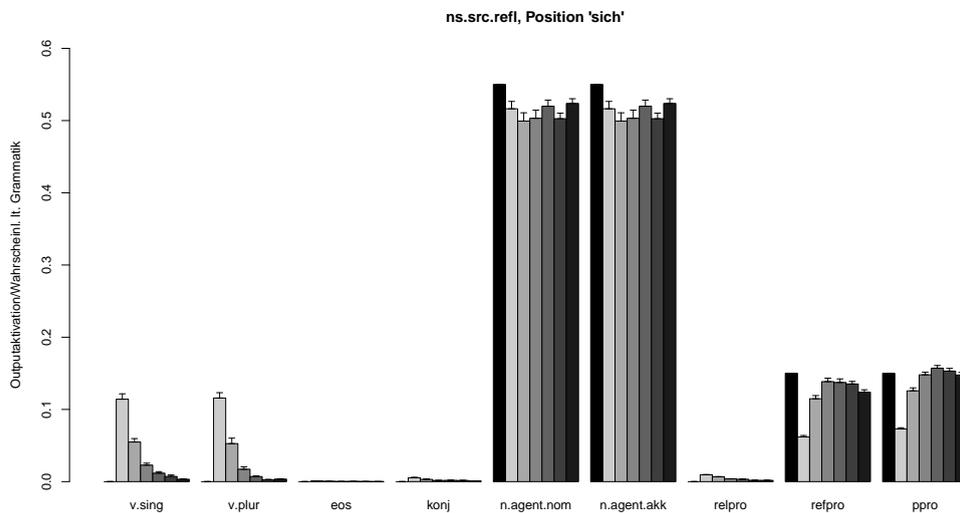


Abbildung 6.5.: Simulation 1. Details der Aktivierung im Vergleich zur absoluten Wortfolge-Wahrscheinlichkeit in ihrer Entwicklung über die Epochen für die Position der Reflexiv- und Personalpronomen im Nebensatz nach eingebettetem Relativsatz (*Als Daniel der Sascha sieht sich kratzt ...*).

6.4) beruht auf einer falschen Aktivierung von Verben mit falscher Kongruenz sowie (neben einigen anderen) vom EOS-Marker und den Konjunktionen. Dies entspricht einem False-Alerts-Fehler für LSK, wie er von Konieczny, Ruh und Müller (2009) gezeigt wurde: Nach der lokalen Sequenz *Wibke.akk amuesiert.mask sich* bzw. *Wibke.nom amuesiert.fem sich* wird das Ende des Satzes und folglich ein Punkt oder eine Konjunktion erwartet. Hier sind zwei Bemerkungen hinzuzufügen: (a) Die in den Diagrammen dargestellten Markierungen des Kasus (*.akk* und *.nom*) wurden diesen Abbildungen zur besseren Orientierung künstlich zugewiesen, werden aber im Lexikon nicht unterschieden. Folglich spielen sie keine Rolle für die Kohärenz der lokalen Subsequenz. (b) In den Objekt-Relativsätzen ist das Relativsatz-Verb immer kongruent zum direkt zuvor erscheinenden Relativsatz-Subjekt, während in Subjekt-Relativsätzen das direkt zuvor erscheinende Relativsatz-Objekt nur in der Hälfte der Test-Sätze zufällig die gleiche Kongruenz-Markierung zum Relativsatz-Verb aufweist, da dieses kongruent zum Relativpronomen sein muss. Dennoch zeigt sich kein relevanter Unterschied zwischen den beiden Satztypen (aufgrund der beinahe identischen Aktivierungsverteilung verzichten wir auf die Abbildung für den Objekt-Relativsatz). Die falsche Kongruenz zwischen dem vermeintlichen Satz-Subjekt und dem Hauptverb der lokalen Sequenz scheint das Netzwerk nicht von der LSK-Interpretation abzuhalten. Insgesamt wird der Fehler für das Matrix-Verb bereits nach der dritten Epoche minimiert. Der False-Alerts-Effekt ist also sehr kurzlebig. Zudem erfasst das Netzwerk sehr früh die korrekte Kongruenz und löst damit die Fernabhängigkeit sehr erfolgreich auf. Anders als bei den einfachen Hauptsätzen wird der Fehler auf der Position der Pronomen nach den Relativsätzen vor allem in den frühen Epochen durch die Aktivierung der Verben verursacht (vgl. Abb. 6.5). Dies wäre aber nur dann korrekt, wenn die Relativsätze in einen Hauptsatz eingebettet wären. Das Netzwerk scheint diesen Unterschied der Einbettung noch nicht erfasst zu haben. Ab der dritten Epoche aber lässt sich der bleibend hohe

Fehler auf dieser Position nicht mehr befriedigend erklären, denn hier ist die falsche Aktivierung der Verben auf ein Minimum reduziert. Leider konnten wir keine weitere plausible Fehlerquelle identifizieren. Wir müssen also wie bei den einfachen Hauptsätzen als Ursache Verschiebungen über eine vergleichsweise hohe Anzahl zu aktivierender Units und fehlende Aktivierung auf den Eigennamen annehmen (s.o.).

**Komplexe Hauptsätze** Die nächsten beiden Abbildungen zeigen den globalen Fehler für Hauptsätze mit eingebettetem Subjekt-Relativsatz (Abb. 6.6) und nach eingebettetem Objekt-Relativsatz (Abb. 6.7). Im Unterschied zu den zuvor besprochenen Nebensatzkonstruktionen findet sich für diese beiden Satztypen für Epoche 10 der leichte Anstieg des Fehlers während des Relativsatzes nicht mehr, wohl aber der bleibend erhöhte Fehlerwert auf der Position der Reflexiv- und Personalpronomen. In der Fehlerentwicklung zeigt sich wieder der anfangs hohe Fehler mit konstanter Verbesserung auf dem Matrix-Verb. Zudem zeigt sich ein Unterschied zwischen Subjekt- und Objekt-Relativsätzen: Der anfangs etwas erhöhte Fehler auf dem Subjekt bzw. Objekt sowie auf dem Verb in den Relativsätzen wird in Subjekt-Relativsätzen schneller reduziert als in Objekt-Relativsätzen (für einen ähnlichen Effekt vgl. MacDonald & Christiansen, 2002). Ab Epoche 5 gleichen sich die Fehlerwerte einander an.

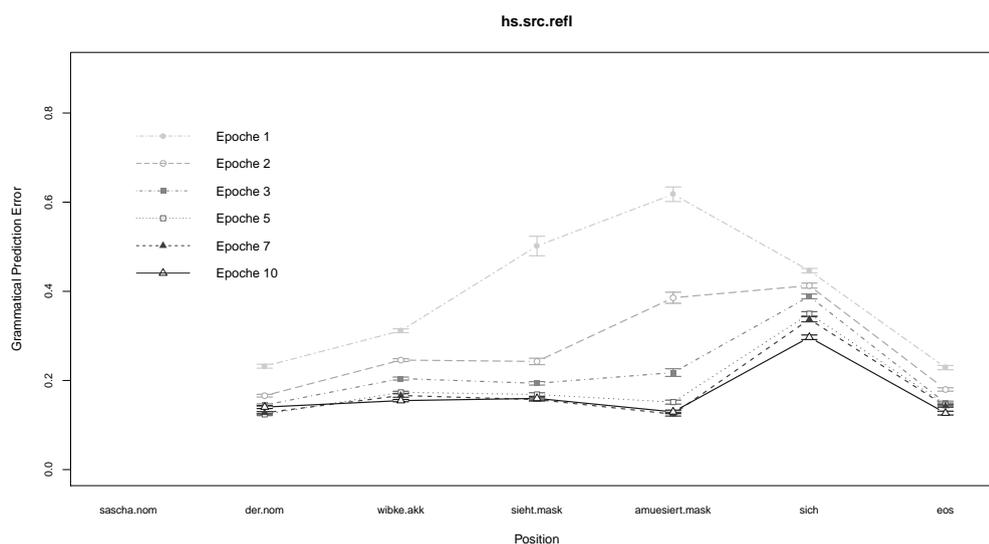


Abbildung 6.6.: Simulation 1. Verlauf und Entwicklung des globalen Fehlers der Wortvorhersage-Komponente, gruppiert nach Epoche: GPEn über Wortpositionen für Hauptsätze mit zentral eingebettetem Subjekt-Relativsatz (Kontrolle 1).

Der leichte Fehler auf dem Objekt im Subjekt-Relativsatz (*wibke.akk*) hat seine Ursache in einer zu starken Aktivierung der korrekten Eigennamen, die auf Kosten der Reflexiv- und Personalpronomen geht. Der parallele Fehler auf dem Subjekt im Objekt-Relativsatz (*wibke.nom*) hingegen resultiert aus einer etwas zu geringen Aktivierung der hier ausschließlich korrekten Eigennamen und einer leichten Aktivierung der Reflexiv- und Personalpronomen. Trotz dieses scheinbaren Unterschieds haben diese Fehler vermutlich die gleiche Ursache: Das Netzwerk kann in den frühen Epochen noch nicht in ausreichendem Maße zwischen

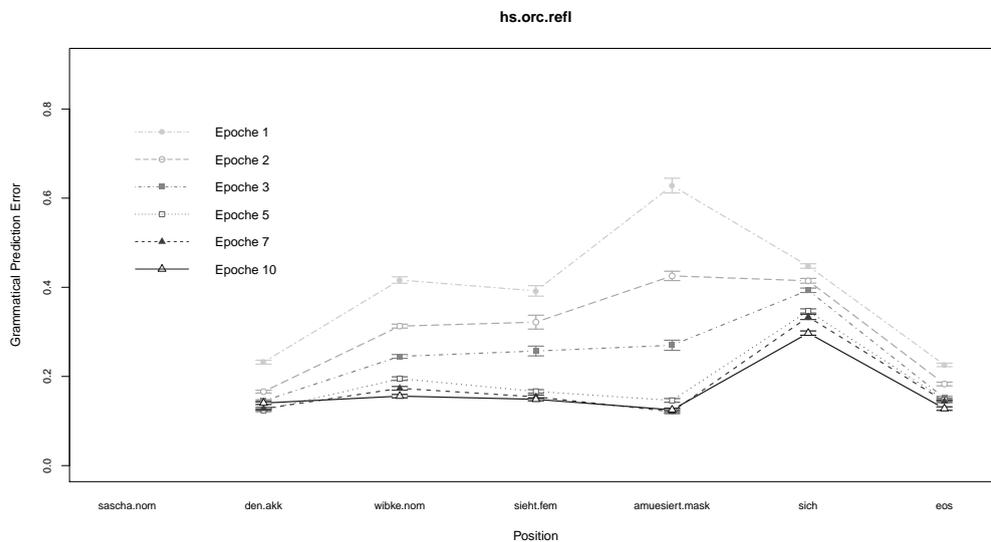


Abbildung 6.7.: Simulation 1. Verlauf und Entwicklung des globalen Fehlers der Wortvorhersage-Komponente, gruppiert nach Epoche: GPEn über Wortpositionen für Hauptsätze mit zentral eingebettetem Objekt-Relativsatz (Kontrolle 2).

Objekt- und Subjekt-Relativsätzen bzw. zwischen den Relativpronomen, die diese einleitenden, unterscheiden. Der Fehler auf dem Relativsatz-Verb im Subjekt-Relativsatz beruht wieder auf einer leichten Aktivierung von Verben mit falscher Kongruenz sowie der zu geringen Aktivierung der hier möglichen Relativpronomen. Im Objekt-Relativsatz hingegen beruht er in erster Linie auf der Aktivierung von Verben mit falscher Kongruenz, die erst mit der fünften Epoche abnimmt. Dies ist erstaunlich, da die Subjekt-Prädikat-Kongruenz bei der minimalen Distanz von einer Position zum bestimmenden Subjekt einfacher sein sollte. Dies ist vermutlich wiederum auf die mangelnde Differenzierung zwischen den beiden Relativsatztypen zurückzuführen, durch die der Interpretation von MacDonald und Christiansen (2002) folgend sich die nicht-kanonische Wortstellung auf den Objekt-Relativsatz nachteilig auswirkt. Eine weitere Beobachtung ist zu diskutieren: In beiden Relativsatztypen beruht der Fehler für das Verb auch auf einer falschen Aktivierung von EOS und Konjunktionen in Epoche 1. Dies ist wiederum ein Anzeichen für einen LSK-Effekt, der bereits von Konieczny, Ruh und Müller (2009) für diese Position beschrieben wurde (vgl. auch Konieczny & Müller, 2010). Zwar fehlt hier in der vorausgehenden lokalen Sequenz *Sascha der Wibke ...* das Verb. Die folgende Fehlaktivierung beruht aber auf einer mangelnden Differenzierung zwischen Relativpronomen und Verben in den frühen Epochen, die auf der Ähnlichkeit ihrer Kollokationen (Positionierung zwischen zwei Nomen) beruht – das Relativpronomen *der* in unserem Beispiel wird also eingeschränkt als Verb interpretiert. Diese mangelnde Differenzierung belegen Konieczny und Kollegen anhand multidimensionaler Skalierung, über die sich in den frühen Epochen eine räumliche Zuordnung der Relativpronomen zu den Verben zeigen lässt (vgl. Kap. 3.2.3).<sup>29</sup>

<sup>29</sup>Analysen mit explorativer Faktorenanalyse weisen in unseren Daten teilweise ebenfalls auf diese Ursache hin. Allerdings konnten wir eine solche gemeinsame räumliche Untergliederung von Relativpronomen und Verben nur für einzelne Netzwerke finden. Dieser Effekt ist also sehr instabil und legt die Annahme nahe, dass der hieraus resultierende LSK-Effekt nur auf unzureichendem Training beruht.

Der Fehler auf dem Matrix-Verb (*amüsiert.mask*) während der ersten Epochen beruht im Falle beider Relativsatztypen in erster Linie auf einer Aktivierung von Verben mit falscher Kongruenz. Hinzu kommt eine geringfügige Aktivierung der Eigennamen in der ersten Epoche, die auf eine Interpretation der lokalen Subsequenz (*Wibke sieht ...*) als Hauptsatzstruktur hinweist. Der Fehler auf der Position der Reflexiv- und Personalpronomen beruht während der ersten Epochen wieder auf einer zu geringen Aktivierung dieser pronominalen Formen und einer anfangs zu starken Aktivierung der ebenfalls korrekten Eigennamen.

### 6.3.3.2. Performanz bei der Referenzzuweisung

Auch für die Performanz bei der Referenzzuweisung zeigt sich schon in frühen Epochen ein sehr niedriger Fehler. Dieser ist allerdings darauf zurückzuführen, dass der zu aktivierende Referent für die meisten Input-Wörter als primitive Eins-zu-eins-Assoziation sehr einfach zu ermitteln ist.<sup>30</sup> Die einzigen Positionen, auf denen keine eindeutige bzw. nur eine kontextabhängige Referenzzuweisung möglich ist, sind die Relativ-, Reflexiv- und Personalpronomen. Für diese zeigt sich auch noch nach einer höheren Anzahl von Epochen eine Verbesserung des Fehlers. Nach ca. 60 Epochen verbessert sich die Zuweisungsgenauigkeit nur noch geringfügig. Das arithmetische Mittel für die Summe aller Zuweisungsfehler beläuft sich nach 60 Epochen für den allgemeinen Fehler auf  $\bar{x}_{\text{RAEn}} = 0.013$  ( $\sigma = 0.054$ ), für alle pronominalen Wortformen auf  $\bar{x}_{\text{RAEn}} = 0.059$  ( $\sigma = 0.102$ ), und bei der ausschließlichen Berücksichtigung der Reflexiv- und Personalpronomen auf  $\bar{x}_{\text{RAEn}} = 0.111$  ( $\sigma = 0.118$ ).

Der globale Fehler RAE ist wenig aufschlussreich, da der Fehlerwert bereits ab den ersten Epochen für praktisch alle Satzpositionen minimal wird. Dies schließt auch die Relativpronomen ein, die direkt auf ihr Antezedens folgen und dadurch problemlos zu bestimmen sind. Wir verzichten deshalb auf eine Darstellung im Diagramm. Die einzige interessante Position sind die Reflexiv- und Personalpronomen, für die das Netzwerk erst Strukturwissen erwerben muss und für die der Fehler nur im Falle der Reflexivpronomen in einfachen Hauptsätzen minimal wird. Die Verläufe der Verbesserung des Fehlers stellen sich für die anderen Satzstrukturen sehr ähnlich dar, sodass die Unterschiede über die einzelnen Verläufe hinweg kaum sichtbar werden. Deshalb erfassen wir den Verlauf der Fehlerentwicklung auf den Reflexiv- und Personalpronomen und den Unterschied zwischen den Satzstrukturen anhand übergreifender Liniendiagramme für die beiden pronominalen Formen (Abb. 6.8 und 6.9). Die Datenpunkte geben das Mittel des RAE der pronominalen Formen gruppiert nach den Satztypen des Test-Sets im Verlauf nach den Epochen in 10er-Schritten bis zur 100. Epoche wieder. Die Fehlerbalken zeigen den Standardfehler.<sup>31</sup>

---

<sup>30</sup>Für alle definiten Nomen ergibt sich eine feste kontextunabhängige Eins-zu-eins-Referenzzuweisung zu ihrem Referenten, für alle anderen nicht auf Objekt-Referenz bezogenen Wörter eine noch einfachere Zuweisung zum Non-Reference-Knoten. So erklärt sich der niedrige allgemeine Fehlerwert nach der 10. Epoche von  $\bar{x}_{\text{RAEn}} = 0.046$  ( $\sigma = 0.154$ ). Diesem steht ein sehr hoher auf Anaphern bezogener Fehlerwert gegenüber: Für alle pronominalen Formen beträgt  $\bar{x}_{\text{RAEn}} = 0.201$  ( $\sigma = 0.274$ ). Wenn ausschließlich Reflexiv- und Personalpronomen berücksichtigt werden, erhöht sich der Mittelwert noch auf  $\bar{x}_{\text{RAEn}} = 0.378$  ( $\sigma = 0.275$ ).

<sup>31</sup>Die Legende unterscheidet die Satztypen nach *HS* (Hauptsatz), *NS* (Nebensatz), *SRC* (eingebetteter Subjekt-Relativsatz) und *ORC* (eingebetteter Objekt-Relativsatz).

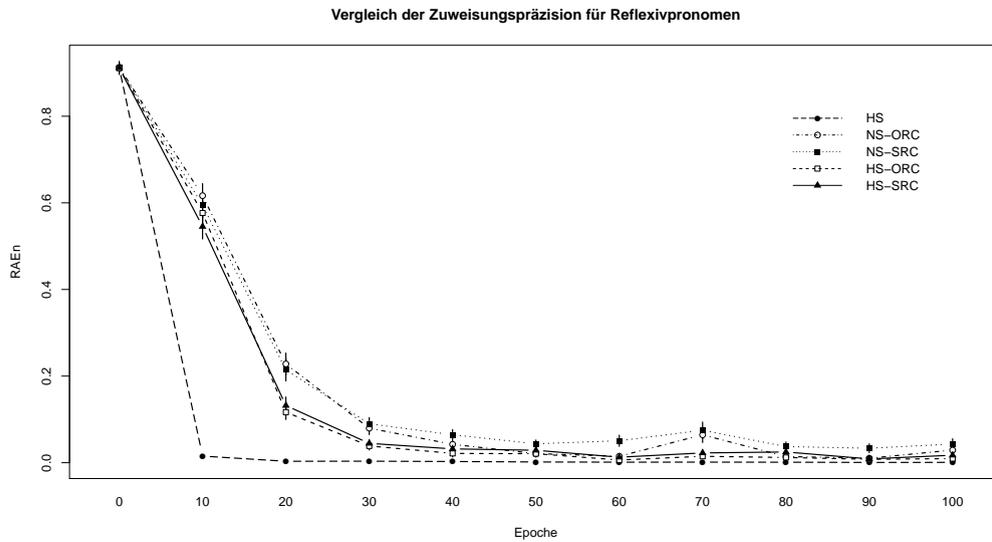


Abbildung 6.8.: Simulation 1. Verlauf und Entwicklung des globalen Fehlers für die Referenzzuweisungs-Komponente: RAEn für Reflexivpronomen nach Satztypen über Epochen.

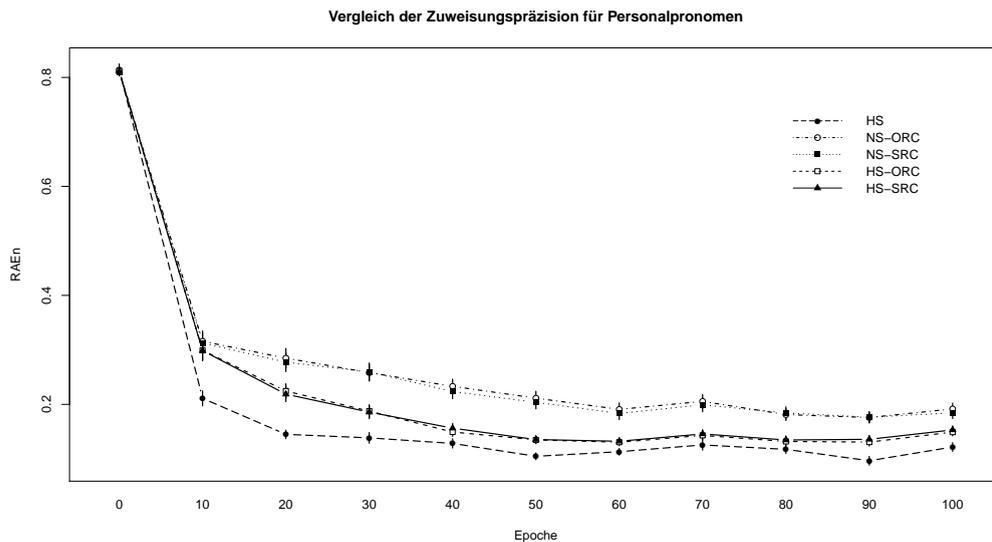


Abbildung 6.9.: Simulation 1. Verlauf und Entwicklung des globalen Fehlers für die Referenzzuweisungs-Komponente: RAEn für Personalpronomen nach Satztypen über Epochen.

**Globale Analyse** In den Abbildungen 6.8 für die Reflexivpronomen und 6.9 für die Personalpronomen<sup>32</sup> zeigen sich folgende Merkmale: Zwischen den pronominalen Formen sind die Anfangsfehler und die weiteren Entwicklungs- bzw. Verbesserungsverläufe sehr unterschiedlich. Die Reflexivpronomen zeigen einen sehr hohen Anfangsfehler, der sich im Vergleich zum Personalpronomen wesentlich schneller verringert und sich ab Epoche 50 auf einem geringeren Fehlerwert als das Personalpronomen stabilisiert. Der Fehler für die Personalpronomen ist anfangs niedriger, nimmt aber wesentlich langsamer ab und stabilisiert sich ab etwa Epoche 60 auf einem höheren Fehlerwert als bei den Reflexivpronomen. Innerhalb beider Formen zeigen sich Unterschiede zwischen den Satztypen, wobei sich hinsichtlich ähnlicher Werte und Verläufe drei Gruppierungen unterscheiden lassen: (a) Hauptsätze, (b) Hauptsätze mit Relativsatz und (c) Nebensätze mit Relativsatz. Die Art des eingebetteten Relativsatzes hingegen scheint keine Performanz-Unterschiede zu erzeugen. Bei den Personalpronomen zeigen die Pronomen nach Relativsatz in Nebensätzen durchgehend die höchsten Fehlerwerte. Danach folgen Pronomen nach Relativsatz in Hauptsätzen, die geringsten Fehlerwerte finden sich in einfachen Hauptsätzen. Diese Rangfolge bleibt über alle Epochen hinweg erhalten. Bei den Reflexivpronomen etabliert sich vorübergehend dieselbe Rangfolge. Zudem zeigen sich kurzzeitig geringfügig schlechtere Werte für Nebensätze mit Subjekt-Relativsatz. Anders als bei den Personalpronomen minimieren sich die Unterschiede für die verschiedenen Satztypen nach Epoche 40, eine Rangfolge lässt sich nicht mehr erkennen.

Woher stammen diese Unterschiede einerseits zwischen den Satztypen und andererseits zwischen den Anapherntypen? Wir ziehen dazu wieder die Verteilungen der Output-Aktivierung heran. Anders als für die Wortvorhersage-Komponente können wir bei der Referenzzuweisungs-Komponente aus der Darstellung der Output-Aktivierung über mehrere verschiedene Sätze keine sinnvollen Ergebnisse ableiten, da wir daraus nicht die Verteilung über verschiedene Wortklassen erkennen können, sondern nur die Verteilung über die verschiedenen Referenten. Damit würde über mehrere Sätze gemittelt jeglicher mögliche Effekt nivelliert werden, der ja gerade im Verhältnis der kontextabhängigen Aktivierung der Referenten erwartet wird. Anhand der Aktivierungsverteilungen einzelner Sätze hingegen lässt sich nur beschränkt eine sinnvolle Aussage treffen. Aus diesem Grund haben wir eine erweiterte Analyse durchgeführt, in der die Aktivierungen der verschiedenen Referenten anhand verschiedener Gesichtspunkte Referentenklassen zugewiesen werden. Über diese Klassen lassen sich Aktivierungsverteilungen über mehrere Sätze und Netzwerke sinnvoll vergleichen.

**Detailanalyse: Reflexivpronomen** Wir erläutern die Klassen anhand eines Beispielsatzes (entsprechend Bsp. 22b): *Als Daniel<sub>i</sub> den Sascha<sub>j</sub> sieht sich<sub>i/\*j</sub> kratzt langweilt Sarah Wibke*. Für Reflexivpronomen werden folgende Referentenklassen unterschieden:

- Antezedens (1\_Antezedens): Subjekt des Matrixsatzes und damit korrektes Antezedens des Reflexivpronomens (*Daniel<sub>i</sub>*).
- Genus-kongruenter Distraktor (2\_DistraktorKongruent): Referent des im Relativsatz genannten Eigennamens und damit direkt vor dem Pronomen genanntes Subjekt oder

---

<sup>32</sup>Wir geben nur die Version mit maskulinem Personalpronomen wieder, die Abbildungen für Sätze mit femininem Personalpronomen zeigen ein identisches Bild der Verläufe und Relationen.

Objekt. Allerdings werden in dieser Klasse nur diejenigen Referenten berücksichtigt, deren Genus mit dem des jeweils korrekten Antezedens übereinstimmt (*Sascha*).

- Nicht genus-kongruenter Distraktor (3\_DistraktorNichtK): Nochmals der Referent des im Relativsatz genannten Eigennamens. Allerdings werden in dieser Klasse alle Referenten berücksichtigt, unabhängig von der Übereinstimmung zum Genus des korrekten Antezedens (im Beispielsatz wiederum *Sascha*; aber auch die femininen Referenten würden die Voraussetzungen erfüllen, wenn sie auf dieser Position stehen).<sup>33</sup>
- Genus-kongruente Referenten (4\_GenusKongruent): Mittelwert der Aktivierung aller Referenten, deren Genus mit dem des korrekten Antezedens übereinstimmt. Es werden alle Referenten mit entsprechendem Genus berücksichtigt außer das Antezedens selbst (für unseren Beispielsatz also die maskulinen Referenten *Lars* und *Sascha*).
- Nicht genus-kongruente Referenten (5\_GenusNichtK): Mittelwert der Aktivierung aller Referenten, deren Genus nicht mit dem des korrekten Antezedens übereinstimmt (*Bettina*, *Wibke* und *Sarah*).

Die Diagramme in Abbildung 6.10 zeigen die Aktivierungsverteilung über die Referentenklassen für Reflexivpronomen für die Epochen 10 bis 40 (d.h. die Epochen, auf denen sich im globalen Fehler die deutlichsten Unterschiede zeigen). Beginnen wir mit den nicht genus-kongruenten Referenten und Distraktoren: Referenten dieser beiden Klassen werden von Beginn an nicht aktiviert (vgl. Epoche 10). Die Netzwerke können also frühzeitig Informationen zum korrekten Genus aufrechterhalten, intervenierende Referenten mit abweichendem Genus nehmen keinen Einfluss auf die Anaphernauflösung. Interessanter ist die Klasse des korrekten Antezedens: Einfache Hauptsätze (HS) unterscheiden sich mit einer korrekten Auflösung ab Epoche 10 deutlich von den komplexeren Satztypen. Zudem zeigt sich ein Unterschied zwischen Nebensätzen (NS-SRC und NS-ORC) und Hauptsätzen mit eingebettetem Relativsatz (HS-SRC und HS-ORC). Bis Epoche 30 zeigen die Netzwerke die schlechteste Performanz für die Satztypen mit vorangestelltem Nebensatz. Während wir also eine Rangfolge Hauptsatz – Hauptsatz mit Relativsatz – vorangestellter Nebensatz mit Relativsatz finden, scheint der Typ des Relativsatzes innerhalb eines Matrixsatztyps keinen Unterschied zu machen. Ab Epoche 40 ist die Rangfolge nahezu nivelliert. Die genau entgegengesetzte Rangfolge finden wir in etwas geringerer Ausprägung für genus-kongruente Referenten und Distraktoren: In den Nebensatztypen werden diese etwas stärker aktiviert als in Hauptsätzen mit Relativsatz, in einfachen Hauptsätzen spielen genus-kongruente Referenten keine Rolle. Die Tatsache, dass sich der Unterschied in den unabhängig von ihrer Nennung falsch aktivierten Referenten im gleichen Maße abbildet wie bei den zuvor genannten Referenten, weist darauf hin, dass die Interferenz-Effekte eher auf einer generellen genusabhängigen Fehlaktivierung beruhen und nicht auf einem Effekt der genannten Distraktoren. Zudem scheint in Epoche 10 die gemittelte Aktivierung aller genus-kongruenten Referenten höher zu sein als die des Distraktors. Dieser scheint also aufgrund seiner syntaktischen Position vorübergehend als unwahrscheinlicher betrachtet zu werden.

<sup>33</sup>Da in einfachen Hauptsätzen kein intervenierender Referent genannt wird, werden bei diesem Satztyp auch keine Werte zugewiesen. Entsprechend kann hierfür auch kein Vergleich angestellt werden.

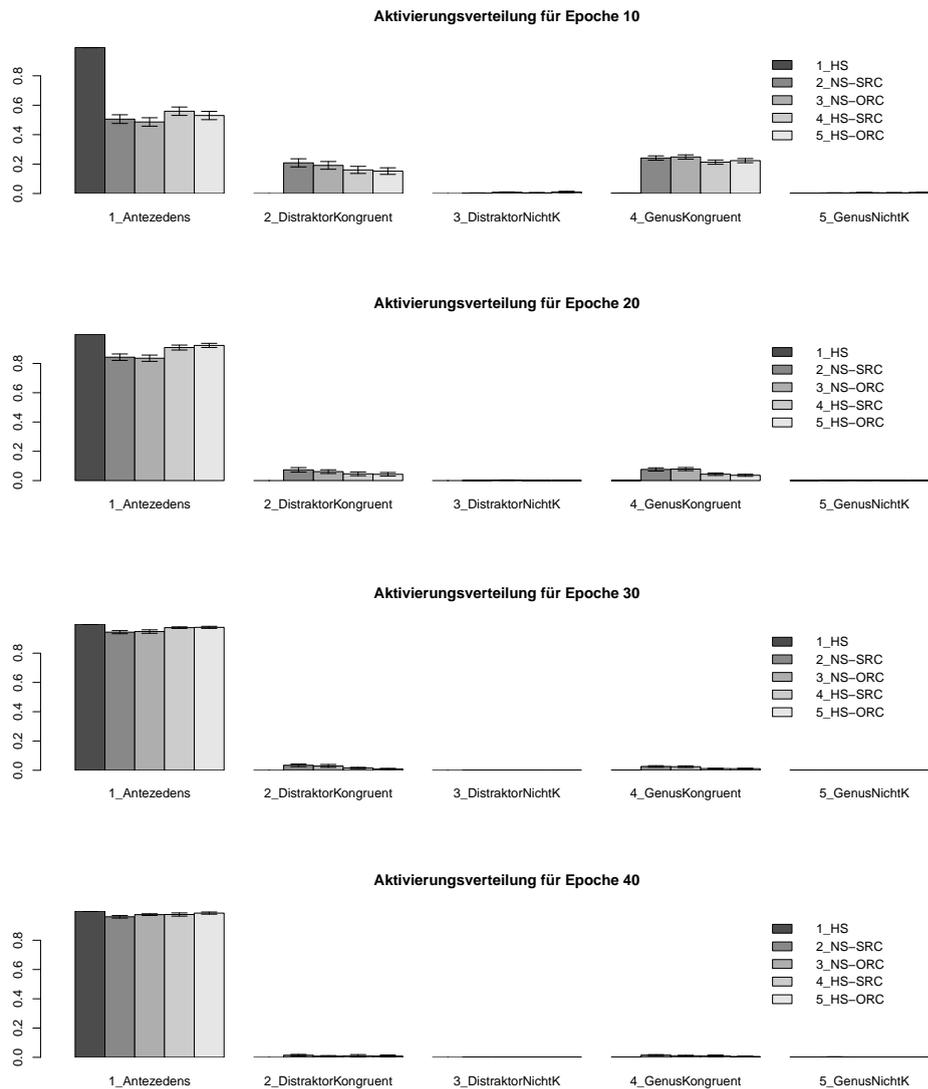


Abbildung 6.10.: Simulation 1. Details der Referenten-Aktivierung für Reflexivpronomen über die Epochen 10 bis 40. Die Balken zeigen die Aktivierung für die verschiedenen Satztypen gruppiert nach Referentenklasse (Maximalwert für Wahrscheinlichkeiten und Aktivierungen ist 1.0).

Um die benannten Unterschiede zu überprüfen, haben wir paarweise vergleichende t-Tests gerechnet. Diese ermöglichen einen genaueren Einblick in die Daten und helfen bei der Entscheidung, welche Unterschiede in den Diagrammen als Effekt interpretiert werden dürfen. Wir verwenden dafür die Funktion *pairwise.t.test* aus dem Standard-R-Package *stats* (R Development Core Team, 2010). Zur Korrektur der p-Werte für mehrfache Tests verwenden wir die Methode nach (Holm, 1979, Bonferroni-Korrektur ergab in unseren Analysen ähnliche Ergebnisse).<sup>34</sup> Beginnen wir mit dem korrekten Antezedens: Der Unterschied zwischen einfachen Hauptsätzen und komplexen Satztypen ist bis Epoche 20 signifikant für den Vergleich mit allen komplexen Satztypen, für Epoche 30 aber nur noch im Vergleich mit den vorangestellten Nebensätzen. Der Unterschied zwischen vorangestellten Nebensätzen mit Relativsatz und Hauptsätzen mit Relativsatz wird hingegen nur in Epoche 20 signifikant. Der Unterschied zwischen genus-kongruenten Distraktoren wird für keine Epoche signifikant. Ähnlich wie für die Rangfolge beim korrekten Antezedens sieht es in der Klasse der genus-kongruenten Referenten aus: In Epoche 10 unterscheiden sich die einfachen Hauptsätze noch signifikant von allen komplexen Satztypen, in Epoche 20 und 30 nur noch von den vorangestellten Nebensätzen. Zwischen Nebensätzen und Hauptsätzen mit Relativsatz werden nur die Unterschiede in Epoche 20 signifikant. Signifikante Unterschiede ergeben sich insgesamt betrachtet immer nur zwischen den Matrixsatztypen, nicht aber zwischen den Relativsatztypen innerhalb der Matrixsätze. Der für Epoche 10 ermittelte Unterschied zwischen genus-kongruenten Referenten und Distraktoren wird tatsächlich zugunsten der allgemeinen Referenten-Aktivierung signifikant.

**Detailanalyse: Personalpronomen** Für Personalpronomen gestalten sich die Referentenklassen etwas anders, da hier keine Referenten zwingend als Antezedens vorgegeben sind, sondern nur anhand syntaktischer (Subjekt des Matrixsatzes) und morphosyntaktischer Kriterien (hier dem Genus) als Koreferenten ausgeschlossen werden. Wir erläutern die Klassen wieder anhand unseres Beispielsatzes (entsprechend Bsp. 22b): *Als Daniel<sub>i</sub> den Sascha<sub>j</sub> sieht ihn\*<sub>ij</sub> kratzt langweilt Sarah Wibke*. Es werden folgende Referentenklassen unterschieden:

- Antezedenzien (1\_Antezedens): Korrektes Antezedens und weitere mögliche Antezedenzien des Personalpronomens nach syntaktischen und morphosyntaktischen Bedingungen (im Beispielsatz gleichermaßen *Lars* und *Sascha*, aber nicht *Daniel*).
- Distraktor (2\_Distraktor): Referent des im Relativsatz genannten Eigennamens (*Sascha<sub>j</sub>*). Anders als beim Reflexivpronomen kann dieser bei Genus-Kongruenz zum Personalpronomen valides Antezedens des Pronomens sein.
- Subjekt des Matrixsatzes (3\_Matrixsubjekt): Als Antezedens zwingend ausgeschlossenes Subjekt des Matrixsatzes (*Daniel<sub>i</sub>*).

Die Diagramme in Abbildung 6.11 zeigen die Aktivierungsverteilung über die Referentenklassen für Personalpronomen für die Epochen 10 bis 80. Wir ergänzen die Vergleiche jeweils um den Befund, welche Unterschiede im paarweise vergleichenden t-Test mit Korrektur nach Holm signifikant wurden. Beginnen wir mit dem Antezedens: In Epoche

<sup>34</sup>Leider werden für die Funktion *pairwise.t.test* nur p-Werte ausgegeben. Wir verzichten auf den Bericht dieser p-Werte und geben stattdessen nur an, ob diese signifikant wurden (d.h.  $p < 0.05$ , wobei in den allermeisten Fällen wesentlich geringere p-Werte erreicht wurden).

10 unterschieden sich einfache Hauptsätze signifikant von komplexen Satztypen, diese untereinander aber nicht. Der Unterschied zu den komplexen Satztypen bleibt bis Epoche 30 erhalten. Ab Epoche 40 wird der Unterschied nur noch zu den vorangestellten Nebensätzen signifikant. Dieser Unterschied geht erst in Epoche 80 verloren. Ein Unterschied zwischen den komplexen Satztypen stellt sich ab Epoche 20 ein, und hält sich bis Epoche 70: Die korrekten Antezedenzen werden in vorangestellten Nebensätzen signifikant schwächer aktiviert als in Hauptsätzen mit Relativsatz. Für den Distraktor zeigt sich ein ganz ähnliches und über die Epochen noch konsistenteres Bild: Von Epoche 10 bis Epoche 80 werden korrekte Antezedenzen signifikant schwächer aktiviert, wenn sie nach dem Relativsatz in vorangestellten Nebensätzen erscheinen, als wenn sie nach dem Relativsatz in Hauptsätzen erscheinen (in Epoche 80 für NS-ORC nur als Tendenz). Für die Klasse der Referenten, die im jeweiligen Satz durch das Matrixsatz-Subjekt benannt werden, ergibt sich die entsprechend genau umgekehrte Rangfolge: Die falsche Aktivierung des Matrixsatz-Subjekts findet sich am geringsten bei einfachen Hauptsätzen, etwas stärker bei den Hauptsätzen mit Relativsatz und am stärksten bei den vorangestellten Nebensätzen mit Relativsatz. Die Unterschiede dieser Rangfolge sind von Epoche 20 bis 40 signifikant (in Epoche 40 zwischen HS und HS-ORC nur noch als Tendenz). Auf Epoche 10 unterscheiden sich nur einfache von komplexen Sätzen. Ab Epoche 50 wird der Unterschied zwischen einfachen und komplexen Hauptsätzen nicht mehr signifikant. Der Unterschied zwischen Neben- und Hauptsätzen mit Relativsatz bleibt von Epoche 20 bis 80 signifikant. In allen drei Referentenklassen werden zwar Unterschiede zwischen dem Typ des Matrixsatzes signifikant, in keiner der Klassen jedoch findet sich ein Unterschied zwischen Objekt- und Subjekt-Relativsätzen.

### 6.3.3.3. Interne Repräsentation

Die folgenden Diagramme veranschaulichen die Struktur der internen Repräsentationen auf Grundlage der Aktivierungsmuster der Hidden-Layer der beiden Komponenten. Wir untersuchen dabei zum einen die Strukturierung über die Menge aller Wörter, zum anderen vergleichen wir die Struktur für verschiedene Anapherntypen in Abhängigkeit ihres syntaktischen Kontexts (dieses Vorgehen verwenden auch R. Frank et al., 2005). Hierzu verwenden wir als Grundlage das Test-Korpus, das uns eine ausreichend große Anzahl an Sätzen zur Verfügung stellt.<sup>35</sup> Dabei bilden wir für alle distinktiven Einheiten das arithmetische Mittel der repräsentierenden Vektoren über die Vektorenstellen, um einen Datensatz mit jeweils einem Vektor pro distinktiver Einheit zu erhalten, den wir im hochdimensionalen Raum in ein räumliches Verhältnis zu den anderen Einheiten stellen können.<sup>36</sup> Die verschiedenen strukturentdeckenden Verfahren unterscheiden sich in ihren Eigenschaften

---

<sup>35</sup>Im mit 168 Sätzen genügend umfangreichen Daten-Set der Test-Stimuli fehlen nur die Konjunktionen *als* sowie *und*. Ansonsten ist das Test-Korpus mit seiner gleichmäßigen Variation der Verben und Nomen ausgewogen gestaltet.

<sup>36</sup>Für alle Vektoren eines Wortes werden also die einzelnen Vektorstellen verglichen, ihr Mittelwert bildet die entsprechende Vektorstelle des gemeinsamen Mittelwert-Vektors. Dieses Verfahren läuft allerdings Gefahr, durch die Mittelung der Werte Informationen über relevante Gesichtspunkte der räumlichen Gliederung zu verlieren. Möglicherweise werden sogar artifizielle räumliche Eigenschaften abgebildet – eine Problematik, die bereits von Elman (1990) angesprochen wurde. Wir gehen aber davon aus, dass die zentralen Tendenzen der räumlichen Gliederung hinsichtlich der für uns relevanten Eigenschaften in genügender Granularität erfasst werden.

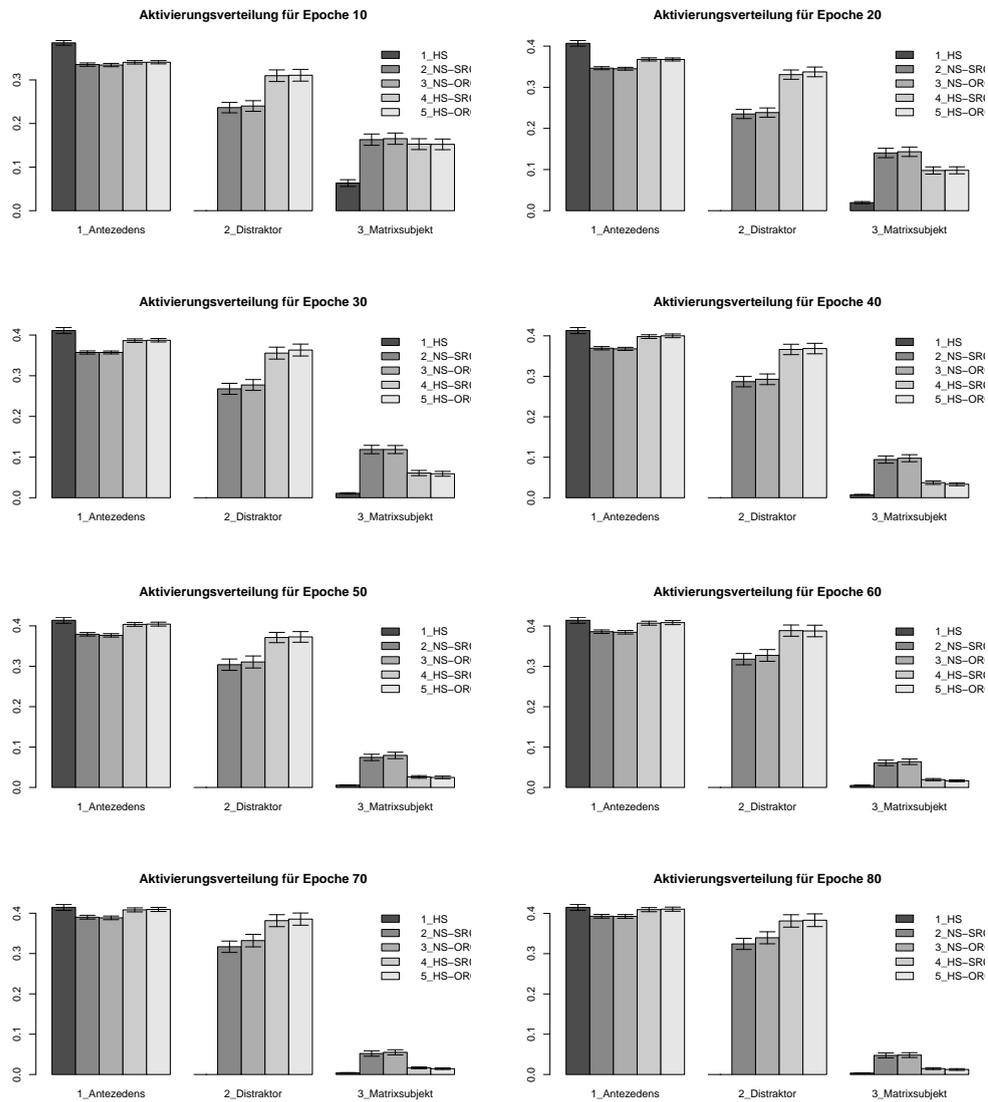


Abbildung 6.11.: Simulation 1. Details der Referenten-Aktivierung für Personalpronomen über die Epochen 10 bis 80. Die Balken zeigen die Aktivierung für die verschiedenen Satztypen gruppiert nach Referentenklasse.

der Strukturerrfassung. In Abhängigkeit von ihren Verfahrensweisen bei der Clusterung oder der Dimensionsreduktion sowie den Algorithmen zur Schätzung von Distanzen können sie zu sehr unterschiedlichen Lösungen für die Clusterung und die Gliederung der Raumstruktur führen (vgl. Kap. 3.1.3.3). Um die im Folgenden vorgestellten Cluster-Lösungen zu validieren, haben wir für alle untersuchten Fälle immer mehrere alternative strukturentdeckende Verfahren verwendet: Hierarchische Cluster-Analyse (HCA) über agglomerative und divisive Verfahren<sup>37</sup>, Hauptkomponenten-Analyse (PCA) und explorative Faktoren-Analyse (eFA). Zudem kamen Vergleiche von Struktur-Lösungen über Samples zum Einsatz.<sup>38</sup> Bei diesen wird das Daten-Set in mehrere Subsets unterteilt, für die jeweils eine eigene Zusammenhangsstruktur ermittelt wird. Diese werden anschließend verglichen und der Überschneidungsgrad ermittelt. Hierfür kam das Verfahren Neighbor-Joining-Tree-Estimation mit Sample-Vergleich über Temperature Nodes sowie über Consensus Clustering zum Einsatz. Wir stellen nicht alle Analysen gesondert vor, sondern konzentrieren uns weitgehend auf Lösungen mit HCA. Auffälligkeiten in anderen Analysen werden erwähnt.

**Kontextunabhängige Analyse aller Wörter** Die ersten beiden Diagramme geben die Struktur der beiden Hidden-Layer wieder, die in Abhängigkeit vom jeweils am Input-Layer dargebotenen Wort entsteht. Gegenstand sind also die Mittelwert-Vektoren aller unterscheidbarer Wörter unabhängig vom Kontext ihres Auftretens. Wir verwenden an dieser Stelle die Daten aus Netzwerken nach 100 Epochen, d.h. nach vollständigem Training.<sup>39</sup> Abbildung 6.12 zeigt die Clusterung des Hidden-Layers der Wortvorhersage-Komponente anhand einer agglomerativen HCA. Die Clusterung zeigt eine saubere Kategorisierung (von links nach rechts) nach Verben und nach nominalen Formen. Die Verben sind untergliedert nach ihrer Kongruenz-Klasse in maskuline und feminine Verben. Die nominalen Formen sind weiter untergliedert nach Relativpronomen gegenüber Nomen sowie Reflexiv- und Personalpronomen. Die Relativpronomen sind nach Genus feingegliedert. Ihre etwas von den Nomen und anderen Pronomen getrennte Gruppierung ist vermutlich auf ihre Kollokationen zurückzuführen, die von denen der anderen nominalen Formen abweicht. Auch die Nomen sind nach Genus feingegliedert. Nahe zu den der Reflexiv- und Personalpronomen sind schließlich die Konjunktionen und der EOS-Marker geclustert. Die Feingliederung der

---

<sup>37</sup>Für die Cluster-Analysen wurden zunächst Korrelations-Matrizen für die Vektorpositionen erstellt. Die Vektorpositionen entsprechen dabei unterscheidbaren Variablen. Da es sich dabei um Aktivierungswerte handelt, zeigt sich über die Daten je Vektorposition eine deutlich linksschiefe Verteilung, weshalb wir als Methode die Spearman-Rangkorrelation verwenden. Die Korrelations-Matrizen wurden in Distanz-Matrizen überführt, die die Grundlage der Cluster-Analyse stellen. Für die agglomerative Clusterung haben wir die Methode *complete linkage clustering* verwendet, als Metrik zur Distanzbemessung in der divisiven Clusterung verwenden wir die euklidische Distanz. Die Clusterlösungen blieben aber auch über andere Clusterungsverfahren konstant.

<sup>38</sup>Für die Analysen mit strukturentdeckenden Verfahren haben wir Funktionen verschiedener R-Packages verwendet: Aus dem Package *stats* (R Development Core Team, 2010) stammen die Funktionen *cor* und *dist*, die Funktion *hclust* für die agglomerative HCA, die Funktion *prcomp* für die PCA sowie die Funktion *factanal* für die explorative Faktorenanalyse. Aus dem Package *cluster* (Maechler, Rousseeuw, Struyf & Hubert, 2005) stammt die Funktion *diana* für die divisive HCA. Aus dem Package *ape* (Paradis, Claude & Strimmer, 2004) verwenden wir die Funktion *nj* für die Neighbor-Joining-Tree-Estimation sowie die Funktion *consensus* für das Consensus Clustering.

<sup>39</sup>Es wurden auch Analysen der internen Repräsentationen für Netzwerke nach Epoche 10 durchgeführt. Die Analysen erbrachten praktisch identische Ergebnisse.

Nomen und der Relativpronomen nach Genus zeigt sich deutlicher in einer PCA oder einer eFA (wir verzichten aber auf eine Abbildung): Hier ist der Zustandsraum horizontal in die verschiedenen Wortformen und vertikal in die verschiedenen Genera unterteilt. Diese Gliederung nach Wortarten und Feingliederung nach grammatischen Merkmalen sollte die Anforderungen für die Wortvorhersage-Aufgabe optimal unterstützen.

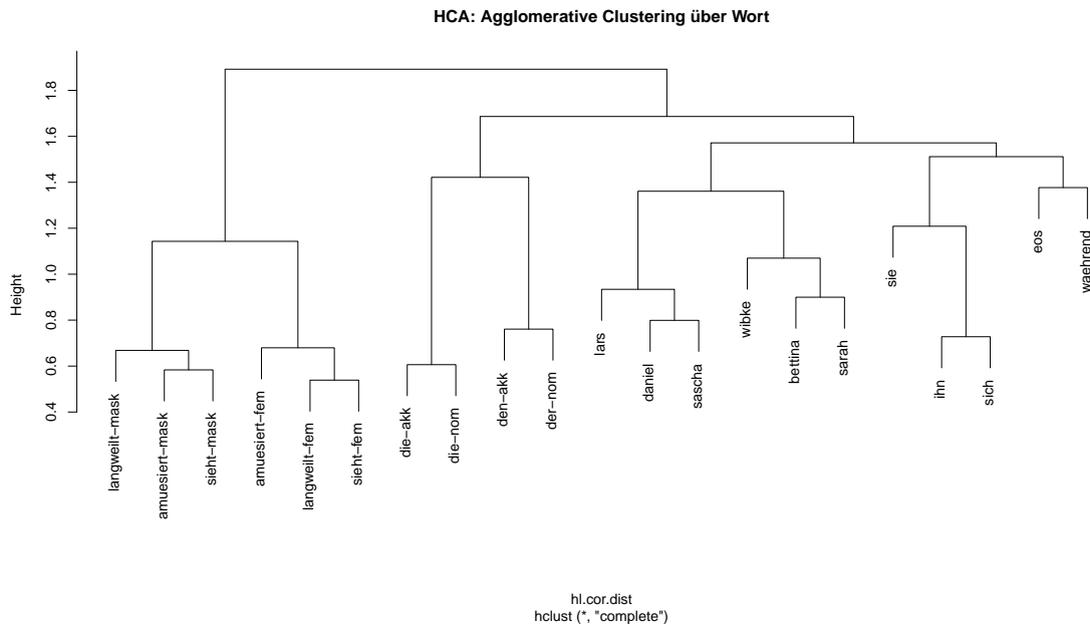


Abbildung 6.12.: Simulation 1. Hierarchische Cluster-Analyse der Hidden-Layer-Aktivierungen der Wortvorhersage-Komponente: Räumliche Gliederung der Wortformen.

Abbildung 6.13 stellt dem eine Analyse des Hidden-Layers der Referenzzuweisungs-Komponente entgegen. In der HCA wird ersichtlich, dass die Untergliederung des internen Zustandsraums nach etwas anderen Kriterien erfolgt als im Hidden-Layer für die Wortvorhersage. Die Untergliederung ist optimiert für die Aufgabe der Referenzzuweisung: Zentrales Clusterungs-Merkmal ist der Genus – über diesen werden alle Wortformen den zwei Haupt-Clustern Femininum und Maskulinum zugewiesen. Dem dritten Haupt-Cluster werden alle Wortformen zugewiesen, die für die Referenzzuweisungs-Aufgabe irrelevant sind, also Verben, Konjunktionen und EOS-Marker, die im Referenz-Output auf den Non-Reference-Marker abbilden. PCA und eFA zeigen zudem die Gewichtung der Informationen: Der dritte Cluster nimmt einen sehr kleinen Bereich im Zustandsraum ein, während die beiden Genus-Cluster den Raum funktional weiter unterteilen. In diesen Analyse-Verfahren wird zudem die Stellung des Reflexivpronomen deutlicher, das in der Cluster-Analyse etwas irreführend den femininen Wortformen zugeordnet ist: In PCA und eFA nimmt diese Proform eine Mittelstellung ein, die der Tatsache gerecht wird, dass es sowohl feminine als auch maskuline Antezedenzen erfassen muss.

**Kontextabhängige Analyse der Anaphern** Die beiden folgenden Diagramme (Abb. 6.14 und 6.15) zeigen die Strukturierung der Hidden-Layer für die Mittelwert-Vektoren der Reflexiv- und Personalpronomen in Abhängigkeit ihres syntaktischen Kontexts. Wir

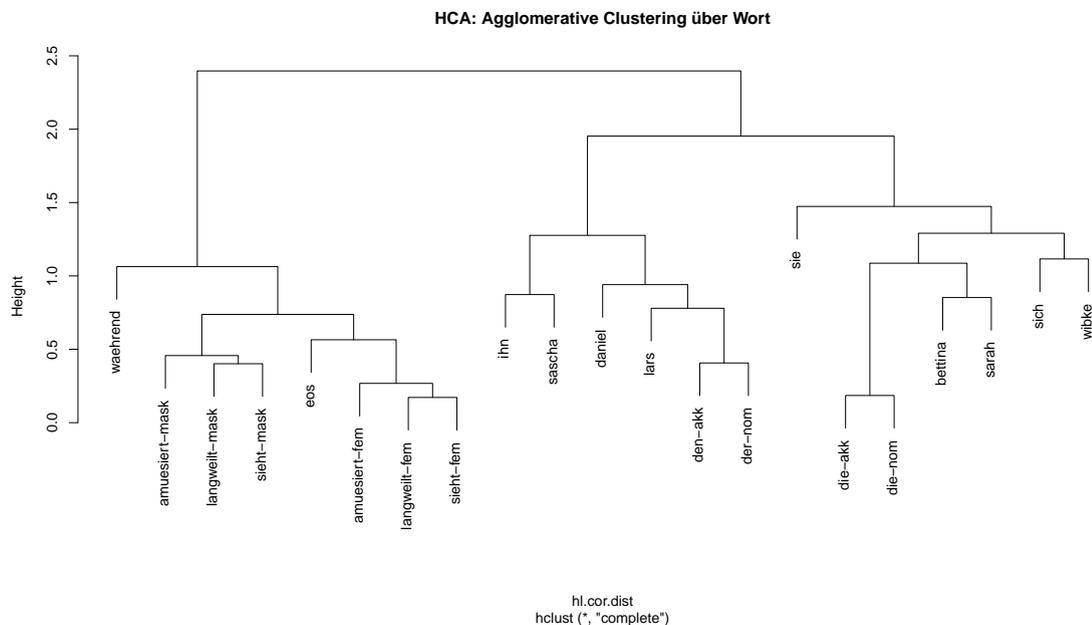


Abbildung 6.13.: Simulation 1. Hierarchische Cluster-Analyse der Hidden-Layer-Aktivierungen der Referenzzuweisungs-Komponente: Räumliche Gliederung der Wortformen.

unterscheiden dabei die Art des Matrixsatzes und des vor dem Pronomen erscheinenden Relativsatzes. Die Mittelwert-Vektoren werden anhand dieser kontextspezifisch unterschiedenen Wörter ermittelt. Die nachstehende Liste führt die Pronomen und die Kürzel für die in den Abbildungen angegebenen Spezifizierungen der Wörter auf:

- Hauptsatz: HS\_
- Nebensatz: NS\_
- Eingebetteter Subjekt-Relativsatz: SRS\_
- Eingebetteter Objekt-Relativsatz: ORS\_
- Pronomen *ihn*: \_ihn
- Pronomen *sie*: \_sie
- Pronomen *sich*: \_sich

Abbildung 6.14 zeigt eine agglomerative HCA für die Wortvorhersage-Komponente. Die Gliederung erfolgt direkt nach der Form des Matrixsatzes: Ein Cluster erfasst die pronominalen Formen in Nebensätzen, der andere Cluster erfasst sie in Hauptsätzen. Interessant ist, dass auf dieser Ebene noch nicht zwischen einfachen und komplexen Hauptsätzen mit eingebettetem Relativsatz unterschieden wird. Die verschiedenen Pronomen werden erst in der Feingliederung unterschieden. Die Untergliederung von eingebettetem Objekt- und Subjekt-Relativsatz fällt auf die niedrigste Ebene und spiegelt damit wider, dass der Inhalt dieser Nebensätze für den weiteren Verlauf des Matrixsatzes syntaktisch und morphosyntaktisch irrelevant ist: Für die Wortvorhersage-Aufgabe ist die Struktur des Matrixsatzes an dieser Stelle wichtig, da durch sie das Folgewort (hier unter anderem

das Pronomen) bestimmt wird, während der abgeschlossene Relativsatz für die folgenden Wortübergangswahrscheinlichkeiten keine Rolle spielt.

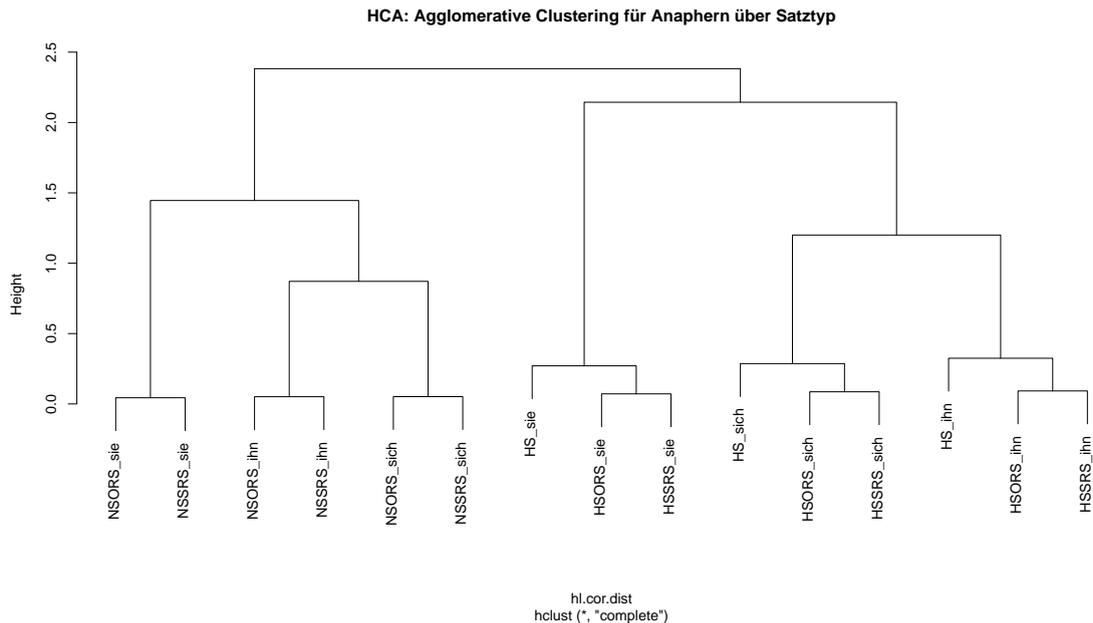


Abbildung 6.14.: Simulation 1. Hierarchische Cluster-Analyse (HCA mit agglomerativem Clusterbildungsverfahren) der Hidden-Layer-Aktivierungen der Wortvorhersage-Komponente: Anaphern in Abhängigkeit ihres syntaktischen Kontexts.

Abbildung 6.15 stellt dieser Struktur eine Clusterung für die Referenzzuweisungs-Komponente gegenüber, die wiederum durch die hier vorgegebene Aufgabe geprägt ist: Die ersten Gliederungsebenen unterscheiden zunächst zwischen den drei pronominalen Formen. Erst auf der jeweils nächsten Ebene wird zwischen verschiedenen Matrixsatztypen unterschieden. Interessant sind zwei Parallelen in den Cluster-Strukturen der beiden verschiedenen Komponenten: zum einen die gemeinsame Gliederung von einfachen und komplexen Hauptsätzen gegenüber Nebensätzen, zum anderen die gemeinsame Gliederung von Reflexiv- und maskulinem Personalpronomen gegenüber den femininen Personalpronomen. Dabei findet sich die Grobgliederung jeweils in der Feingliederung des anderen und umgekehrt. Diese Parallelen beruhen vermutlich darauf, dass sich die Strukturen in Interaktion entwickelt und deshalb wechselseitig aufeinander eingewirkt haben. Etwas überraschend scheint zu sein, dass in Bezug auf die Referenzzuweisungs-Komponente anstatt eines gemeinsamen Clusters für die Personalpronomen eine Clusterung von maskulinen Personal- mit Reflexivpronomen den femininen Personalpronomen gegenübersteht. Die klare Trennung der beiden Personalpronomen ist aber insofern sinnvoll, als diese eine komplementäre Klasse von Referenten bedienen, während das Reflexivpronomen alle Referenten gleichermaßen bedient. Zweidimensionale Abbildungen der räumlichen Gliederung des Zustandsraums durch PCA und eFA (ohne Abb.) klären die Struktur weiter auf: Die Personalpronomen nehmen entgegengesetzte Positionen ein, während die Reflexivpronomen genau zwischen den Personalpronomen positioniert sind, worin sich ihre Unabhängigkeit von der allgemeinen Genus-Markierung widerspiegelt.

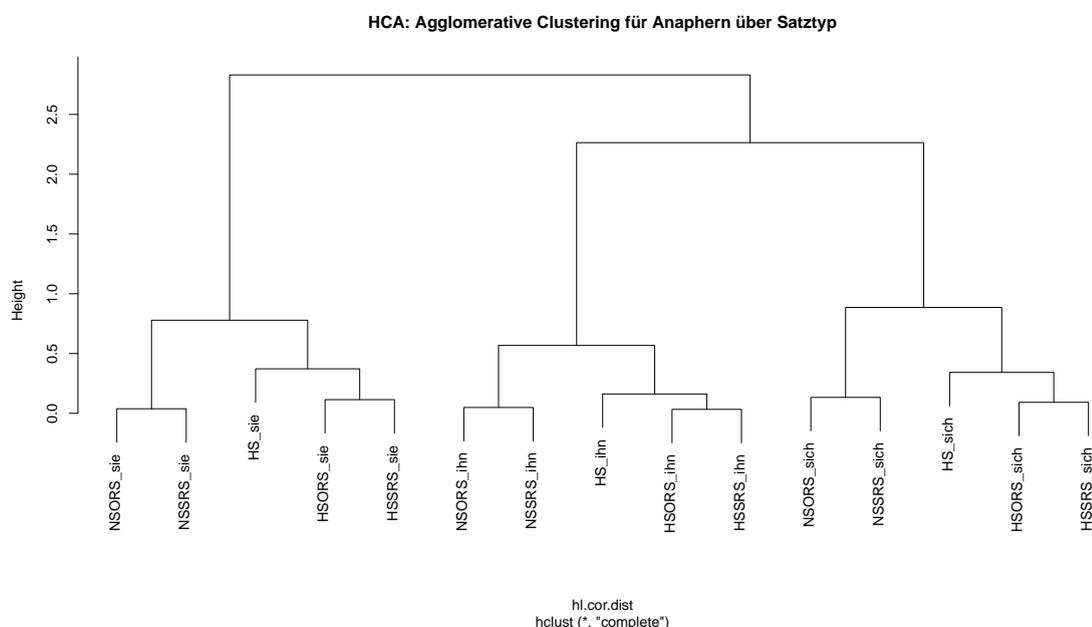


Abbildung 6.15.: Simulation 1. Hierarchische Cluster-Analyse (HCA mit agglomerativem Clusterverfahren) der Hidden-Layer-Aktivierungen der Referenzzuweisungs-Komponente: Anaphern in Abhängigkeit ihres syntaktischen Kontexts.

### 6.3.4. Diskussion

**Wortvorhersage** Generell erbringen die Netzwerke eine in jeder Hinsicht zufriedenstellende Performanz. Die Wortvorhersage-Komponente erbringt bereits in den frühen Epochen Aktivierungsverteilungen, in denen die korrekten Folgewörter am stärksten aktiviert sind. Ab Epoche 10 sind falsche Aktivierungen auf ein Minimum reduziert. Effekte fehlender oder falscher bzw. übermäßiger Aktivierung beschränken sich zum größten Teil auf die ersten 5 Epochen, mitunter sogar nur auf die erste Epoche. Wir haben Effekte gefunden, die bereits in früheren Simulationen gezeigt und in Bezug auf empirische Sachverhalte als Lokalitäts- bzw. Komplexitäts-Effekte, Regularitäts-Effekte und LSK-Effekte beschrieben wurden. Ein allgemeiner Komplexitätsunterschied zwischen Subjekt- und Objekt-Relativsätzen lässt sich aus den Simulationen allerdings kaum ableiten.<sup>40</sup> Für fehlerhafte Aktivierungsverteilungen haben wir in unserer Simulation die folgenden Ursachen identifiziert:

- Effekte mangelnder Differenzierung verschiedener Wortklassen, die durch eine ähnliche Kollokation miteinander verwechselt werden und zu entsprechenden Vorhersagefehlern auf der Folgeposition führen – wie in unseren Daten zum Beispiel beobachtet für das Relativpronomen.

<sup>40</sup>Eventuell kann der leichte Unterschied für Subjekt- und Objekt-Relativsätze in Hauptsätzen in den ersten drei Epochen als Komplexitätsunterschied interpretiert werden. Im Objekt-Relativsatz finden sich hier etwas höhere Werte während des Relativsatzes und eine etwas langsamere Entwicklung des Fehlers auf dem Verb des folgenden Matrixsatzes. Wir halten diese Interpretation aber für sehr spekulativ.

- Effekte mangelnder Differenzierung verschiedener Satzstrukturen, durch die es in die Vorhersage in beiden Satzstrukturen zum wechselseitigen Einbezug möglicher Folgewörter der jeweils anderen Struktur kommt – zum Beispiel beobachtet innerhalb von Subjekt- und Objekt-Relativsätzen.
- Regularitäts-Effekte der Vorhersage von Wortklassen, die auf der gegebenen Position nur in anderen Satzstrukturen erscheinen können, die aber durch ihre höhere Frequenz die Vorhersage der Wortfolgewardrscheinlichkeit in der Satzstruktur mit niedrigerer Frequenz (und entsprechender nicht-kanonischer Wortfolge) beeinflussen – zum Beispiel beobachtet auf den Folgepositionen nach in Nebensätzen eingebetteten Relativsätzen.
- Lokalitäts-Effekte der Reduzierung der Vorhersagequalität für dependente Wortformen über größere Distanzen – zum Beispiel beobachtet im Unterschied zwischen der Vorhersage der Verbform für einfache Hauptsätze und Hauptsätze mit eingebettetem Relativsatz.
- LSK-Effekte: Berücksichtigung lokal kohärenter Subsequenzen in der Wortvorhersage als eigenständige abgeschlossene Strukturen, deren Wortfolgewardrscheinlichkeit mit der global korrekten Folgewardrscheinlichkeit konkurriert – zum Beispiel beobachtet auf dem Matrixsatz-Verb in Nebensätzen nach eingebettetem Relativsatz.
- Distanz-Effekte der Abschwächung syntaktischer Informationen, die nach größeren Distanzen am relevanten Ort nicht mehr verfügbar sind – beschreibbar als eine Mischung aus LSK-, Lokalitäts-, oder Regularitäts-Effekt – zum Beispiel beobachtet für vorangestellte Nebensätze, die am Teilsatz-Ende als normale Nebensätze abgeschlossen werden.

**Referenzzuweisung** Zur Diskussion der Effekte bei der Referenzzuweisung ziehen wir wieder unsere beiden Beispielsätze für die Referentenklassen heran (vgl. Bsp. 23). Bei der Referenzzuweisung werden für Reflexiv- und Personalpronomen ab Epoche 10 korrekte Antezedenzen stärker vorhergesagt als nicht-korrekte Referenten. Allerdings benötigt die Referenzzuweisung für diese pronominalen Formen eine wesentlich längere Trainingszeit als die Wortvorhersage. Relativpronomen stellen hingegen bereits frühzeitig kein Problem dar. Der Unterschied zwischen den Verläufen von Reflexiv- und Personalpronomen lässt sich auf folgenden Umstand zurückführen: Die Auflösung der Reflexivpronomen erfolgt präskriptiv (nur *Daniel* ist möglich), die der Personalpronomen hingegen restriktiv (nur *Daniel* ist ausgeschlossen). Dadurch sind für die Personalpronomen immer mehrere Referenten als Antezedenzen möglich, für die die Berücksichtigung nur weniger lokaler Constraints (hier der Genus des Pronomens) bereits zu einem relativ guten Ergebnis führt. Die Performanz verbessert sich in den folgenden Epochen aber nur sehr langsam, da der restriktive Constraint recht komplex ist. Denn in einem lexikalisch und syntaktisch identischen Kontext können immer mehrere verschiedene Referenten als Antezedens erscheinen. Im Vergleich der Kollokationen über mehrere Sätze hinweg muss das Netzwerk erfassen, dass es immer der Referent des Matrixsatz-Subjekts (*Daniel* in Bsp. 23b) ist, der nicht als Antezedens in Frage kommt. Bei den Reflexivpronomen ist immer nur ein Referent korrektes Antezedens (*Daniel* in Bsp. 23a), wofür grundsätzlich Fernabhängigkeiten aufgelöst und syntaktische Constraints (nur Satz-Subjekte kommen in Frage) berücksichtigt

werden müssen – was zu einem hohen Fehler während der frühen Epochen führt. Da diese Koreferenz aber immer eindeutig zugewiesen wird und keine Variabilität beinhaltet, können die Regelmäßigkeiten ihrer Auflösung vergleichsweise gut erworben werden.

- (23) a. Reflexivpronomen: Als Daniel (*einzig möglicher Antezedens*) den Sascha (*unzulässiger Distraktor*) sieht sich kratzt langweilt Sarah Wibke.
- b. Personalpronomen: Als Daniel (*Matrix-Subjekt, nicht zulässig*) den Sascha (*Distraktor, möglicher Antezedens*) sieht ihn (*ebenfalls möglich: Lars*) kratzt langweilt Sarah Wibke.

Für die Reflexiv- und Personalpronomen lassen sich folgende Quellen für Interferenzen bei ihrer Auflösung identifizieren:

- Distanz- bzw. Lokalitäts-Effekte der Reduzierung der Zuweisungsqualität über größere Distanzen aufgrund verblassender Detail-Informationen.
- Frequenz-Effekte der präziseren Zuweisung in häufigeren Satzstrukturen, die vermutlich nicht auf die Regularität der häufigeren Strukturen zurückgeführt werden können.
- LSK-Effekte: Berücksichtigung lokal kohärenter Subsequenzen für die Referenzzuweisung als eigenständige abgeschlossene Strukturen, deren Constraints mit denen der global korrekten Zuweisung konkurrieren.
- Allgemeine Kontext-Effekte, die sich unabhängig von der syntaktischen Struktur auswirken.

Der Fehler bei den Reflexivpronomen wird in erster Linie durch die ersten beiden Interferenzen bestimmt. Ein LSK-Effekt ist eher auszuschließen, da die Referenten, die unabhängig vom Kontext falsch aktiviert werden, einen gleichen und teilweise größeren Anteil der falschen Aktivierung bedingen als die Referenten im lokalen Kontext. Der lokale Distraktor (in Bsp. 23a *Sascha*) scheint also keine Rolle zu spielen. Die im Vergleich zu den Hauptsätzen schlechtere Performanz für die komplexen Satztypen gegenüber einfachen Hauptsätzen ist in dieser Hinsicht am ehesten durch die abnehmende Präzision der für die Auflösung zur Verfügung stehenden Informationen zu erklären: Über die größere Distanz nimmt die referentenspezifische Information (welcher Referent war Matrixsatz-Subjekt) schneller ab als die allgemeine Genus-Information (welchen Genus hatte das Matrixsatz-Subjekt). Die leichten Differenzen zwischen komplexen Neben- und Hauptsätzen weisen allerdings zusätzlich auf Frequenz-Effekte hin, die zur geringfügig schlechteren Performanz bei den vorangestellten Nebensätzen führen. Die genaue Wortfolge spielt dabei aber keine Rolle für die zu berücksichtigenden Constraints, zumal die Reihenfolge der erscheinenden Referenten in allen komplexen Satztypen identisch ist. Allerdings ist auch unklar, weshalb sich der Frequenzunterschied auf die Auflösung des Reflexivpronomens auswirken sollte.

Die gefundenen Effekte für Personalpronomen lassen sich als LSK-Effekte erklären. Dies zeigt sich am deutlichsten in der Aktivierung des Distraktors (in Bsp. 23b *Sascha*), der als mögliches Antezedens genau gleich stark aktiviert werden müsste wie alle anderen

Referenten mit korrektem Genus, die nicht Subjekt des jeweiligen Satzes sind.<sup>41</sup> Die Distraktoren sind im Falle der vorangestellten Nebensätze am schwächsten aktiviert, da sie hier im lokalen Kontext als vermeintliches Matrixsatz-Subjekt (im Beispielsatz: *Sascha sieht ihn*) als Antezedens tendenziell ausgeschlossen werden. Dies betrifft gleichermaßen die Nebensätze mit Subjekt-Relativsatz, was darauf hinweist, dass die LSK vorwiegend auf der Kohärenz der Wortfolge und unabhängig von weitergehenden grammatischen Informationen wie zum Beispiel abweichender Prädikat-Kongruenz erfolgt. Die korrekte Aktivierung der Klasse der Distraktoren ist in den Bedingungen mit komplexen Hauptsätzen weniger beeinträchtigt, da sie hier nicht in einer lokal kohärenten Subsequenz erscheinen. Dennoch erhält sie auch in diesen Satztypen eine geringere Aktivierung als die Klasse der Antezedenzen. Dies weist darauf hin, dass sich auch der allgemeine Kontext unabhängig von der lokalen syntaktischen Struktur auf diese Reduzierung der Referenten-Aktivierung auswirken kann. Der Effekt der Erfassung des lokalen Referenten (*Sascha*) als vermeintlichem Matrixsatz-Subjekt der lokalen Subsequenz wirkt sich auch wie erwartet auf die Aktivierung des globalen Matrixsatz-Subjekts (*Daniel*) aus: Dieses wird aktiviert, obwohl es als Antezedens ausgeschlossen sein sollte.

Die Ergebnisse der Performanz decken sich teilweise mit den Befunden von R. Frank et al. (2005). Beruhen sie aber tatsächlich auf linearen Generalisierungen und der Unfähigkeit von SRNs, irrelevante lineare Abhängigkeiten zu ignorieren, wie von ihnen angenommen und scheinbar durch Analysen interner Repräsentationen belegt? In unseren Analysen zeigen sich deutlich andere Ergebnisse: Weder im Hidden-Layer der Referenzzuweisungs-Komponente noch im Hidden-Layer der Wortvorhersage-Komponente spielt die Unterscheidung zwischen Objekt- und Subjekt-Relativsatz auf der Position der Anapher eine Rolle. In beiden Komponenten werden hingegen die Hauptmerkmale der Clusterung von Aspekten des syntaktischen Kontexts bestimmt, die für die weiteren Vorhersage- und Zuweisungs-Aufgaben unmittelbar relevant sind – vor allem der Matrixsatztyp und die Form des jeweiligen Pronomens. Irrelevante lineare Generalisierungen lassen sich allerdings nicht feststellen. Die Gliederung für die unterscheidbaren Wörter zeigt zudem eine aufgabenspezifisch unterschiedliche Spezialisierung des Zustandsraums der beiden Komponenten, die die Erfüllung der jeweiligen Aufgabe gewährleistet. Wir gehen davon aus, dass diese Form der Selbstorganisation der internen Strukturen auf unserer Vorgehensweise des holistisch-integrativen Trainings der Netzwerke beruht. Im Umkehrschluss konnten die relevanten Strukturen im Modell von R. Frank et al. (2005) nicht entwickelt werden, da die Aufgaben-Integration unterbunden wurde.

## 6.4. Simulationsstudie 2

In Simulation 2 verwenden wir das wesentlich erweiterte und komplexere Korpus 2. Zum einen geht es wieder um die Untersuchung der Performanz der beiden Komponenten zur Wortvorhersage und zur Referenzzuweisung sowie um die Performanz-Unterschiede für verschiedene Satztypen. Zum anderen wollen wir überprüfen, wie sich die erweiterten

<sup>41</sup>In den Vergleichen kann eine Erklärung der Distraktor-Effekte über die allgemeine Aktivierung aller kongruenten Referenten (wie beim Reflexivpronomen) ausgeschlossen werden, da die Distraktoren ja bereits in der Klasse der möglichen Antezedenzen berücksichtigt werden.

Merkmale wie die genus-unabhängige Subjekt-Prädikat-Kongruenz, die Kasus-Markierung durch Artikel, die deutlich größeren Distanzen zwischen dependenten Elementen, die hinzukommenden Adverbien und nicht zuletzt die satzinterne Interpunktion durch Kommata auf die Performanz auswirken. Wie gut funktioniert die Auflösung von Fernabhängigkeiten der Verb-Kongruenz und der Anaphernbindung in unserer Architektur, wenn dem Korpus eine Grammatik zugrunde liegt, die hinsichtlich Kasusmarkierung, Kongruenzmarkierung und Lexikon näher an einer natürlichen Sprache ist? Darauf basierend wollen wir zudem überprüfen, inwiefern sich die Befunde aus der ersten Simulation für Komplexitätsunterschiede, Lokalitäts-Effekte und LSK-Effekte für Simulationen mit komplexeren Stimuli bestätigen lassen. In den Folgestudien mit den modifizierten Korpora werden wir darüber hinaus untersuchen, welche Performanz-Aspekte der Form der Kongruenzmarkierung zuzuschreiben sind und welche Auswirkungen die erweiterte Interpunktion auf die Performanz hat.

### 6.4.1. Trainings- und Test-Stimuli

Korpus 2 enthält 20.000 Sätze mit insgesamt 326.016 Wörtern (dies ergibt eine durchschnittliche Satzlänge von 16,3 Wörtern). Der längste Satz (einschließlich Satzzeichen) enthält 111 Wörter, der kürzeste 5 Wörter. Das Test-Korpus wurde wiederum manuell nach den grammatischen Vorgaben der PCFG erstellt. Die verwendeten Satzstrukturen werden in Beispiel 24 aufgeführt.<sup>42</sup> Wir verwenden wieder einfache Hauptsätze (Bsp. 24a), in vorangestellte Nebensätze eingebettete Objekt- und Subjekt-Relativsätze (Bsp. 24b, 24c und 24d) sowie in Hauptsätze eingebettete Objekt- und Subjekt-Relativsätze (Bsp. 24e und 24f).

- (24) a. Der Forscher<sup>AGR2</sup><sub>i</sub> kratzt<sup>agr2</sup> sich<sub>i</sub>/ihn<sup>\*</sup><sub>i</sub>/sie<sup>\*</sup><sub>i</sub>.
- b. Als der Forscher<sup>AGR2</sup><sub>i</sub>, den<sub>i</sub> der Bäcker<sup>AGR1</sup><sub>j</sub> sieht<sup>agr1</sup>, sich<sub>i</sub>/<sup>\*</sup><sub>j</sub>/ihn<sup>\*</sup><sub>i</sub>/<sub>j</sub>/sie<sup>\*</sup><sub>i</sub>/<sup>\*</sup><sub>j</sub> kratzt<sup>agr2</sup>, langweilt<sup>agr1</sup> die Soldatin<sup>AGR1</sup> die Ärztin<sup>AGR2</sup>. (*LSK-Bedingung*)
- c. Als der Forscher<sup>AGR2</sup><sub>i</sub>, der<sub>i</sub> den Bäcker<sup>AGR1</sup><sub>j</sub> sieht<sup>agr2</sup>, sich<sub>i</sub>/<sup>\*</sup><sub>j</sub>/ihn<sup>\*</sup><sub>i</sub>/<sub>j</sub>/sie<sup>\*</sup><sub>i</sub>/<sup>\*</sup><sub>j</sub> kratzt<sup>agr2</sup>, langweilt<sup>agr1</sup> die Soldatin<sup>AGR1</sup> die Ärztin<sup>AGR2</sup>.  
(*Kontrollbedingung 1*)
- d. Als der Forscher<sup>AGR2</sup><sub>i</sub>, den<sub>i</sub> der Bäcker<sup>AGR1</sup><sub>j</sub> gerade sieht<sup>agr1</sup>, sich<sub>i</sub>/<sup>\*</sup><sub>j</sub>/ihn<sup>\*</sup><sub>i</sub>/<sub>j</sub>/sie<sup>\*</sup><sub>i</sub>/<sup>\*</sup><sub>j</sub> kratzt<sup>agr2</sup>, langweilt<sup>agr1</sup> die Soldatin<sup>AGR1</sup> die Ärztin<sup>AGR2</sup>.  
(*Kontrollbedingung 2*)
- e. Der Forscher<sup>AGR2</sup><sub>i</sub>, den<sub>i</sub> der Bäcker<sup>AGR1</sup><sub>j</sub> sieht<sup>agr1</sup>, kratzt<sup>agr2</sup> sich<sub>i</sub>/<sup>\*</sup><sub>j</sub>/ihn<sup>\*</sup><sub>i</sub>/<sub>j</sub>/sie<sup>\*</sup><sub>i</sub>/<sup>\*</sup><sub>j</sub>. (*Kontrollbedingung 3*)
- f. Der Forscher<sup>AGR2</sup><sub>i</sub>, der<sub>i</sub> den Bäcker<sup>AGR1</sup><sub>j</sub> sieht<sup>agr2</sup>, kratzt<sup>agr2</sup> sich<sub>i</sub>/<sup>\*</sup><sub>j</sub>/ihn<sup>\*</sup><sub>i</sub>/<sub>j</sub>/sie<sup>\*</sup><sub>i</sub>/<sup>\*</sup><sub>j</sub>. (*Kontrollbedingung 4*)

Anders als in Simulation 1 enthält in Simulation 2 nur noch die Satzstruktur mit in Nebensatz eingebetteten Objekt-Relativsätzen eine lokal kohärente Subsequenz (vgl. Abb. 24b):

---

<sup>42</sup>Wie in Simulation 1 wurde für jeden der abgebildeten Satztypen eine größere Anzahl an Varianten mit verschiedener Auswahl und Reihenfolge der Nomen und Verben erstellt. Satztyp 1 erhält 24 Sätze, die Satztypen 2-6 jeweils 48 Sätze. Insgesamt enthält das Test-Korpus damit 264 Sätze.

*der Bäcker sieht, sich/ihn/sie...* Wir nennen diese Satzstruktur deshalb im Folgenden LSK-Bedingung. Allerdings besteht die LSK nur hinsichtlich der Wortfolge, genau genommen wird sie selbst in der LSK-Bedingung durch das intervenierende Komma unterbunden. Alle anderen komplexen Satzstrukturen unterbinden die LSK über lokal inkohärente Wortfolgen und dienen damit als Kontrollstrukturen. Im vorangestellten Nebensatz mit eingebettetem Subjekt-Relativsatz (Bsp. 24c) sollte der Artikel und die zusätzliche Kasus-Markierung der Nomen eine lokale Lesart mit dem Relativsatz-Objekt als lokalem Subjekt verhindern (Kontrollbedingung 1). In Satz 24d, der wiederum als Nebensatz mit zentral eingebettetem Objekt-Relativsatz realisiert ist, wird die LSK durch ein zusätzlich eingefügtes Adverb unterbrochen (Kontrollbedingung 2): Die Sequenz *der Bäcker gerade sieht, sich/ihn/sie* verliert (zusätzlich zum Komma) durch das intervenierende Adverb *gerade* seine lokal kohärente Struktur. In Satzstruktur 24e ist der Objekt-Relativsatz in einen Hauptsatz eingebettet (Kontrollbedingung 3) und in Satzstruktur 24f ist es ein Subjekt-Relativsatz, der in den Hauptsatz zentral eingebettet ist (Kontrollbedingung 4). In beiden Fällen verhindert das Verb des Matrixsatzes die Kohärenz der Subsequenz (es erscheinen zwei Verben hintereinander). Wie bereits erwähnt, kommt in allen Satzstrukturen noch die Markierung der Satzgrenzen durch Kommata hinzu: Diese markieren den Abschluss des Relativsatzes und unterbrechen damit die lokale Hauptsatz-Struktur zusätzlich. Sollte es bei diesem Korpus dennoch zu LSK-Effekten kommen, erwarten wir diese auf der Position des Kommas nach dem Relativsatz, für das sich dann eine entsprechende alternative Vorhersage im Sinne einer lokalen Lesart ergeben müsste.

#### 6.4.2. Technische Details zur Simulation

Die technischen Details der Simulationen wie Architektur, Lernverfahren und -algorithmen sowie Trainingsparameter decken sich vollständig mit den Angaben, die wir für Simulation 1 dargelegt haben (vgl. Kap. 6.3.2).<sup>43</sup> Der einzige Unterschied betrifft die Größe der Input- und Output-Layer, die sich nach dem Umfang des Lexikons der zugrundeliegenden Grammatik und nach der Menge unterscheidbarer Referenten (zuzüglich eines Non-Reference-Markers) richten. Input- und Output-Layer für die Wortvorhersage erhalten damit 50 Units, der Output-Layer für die Referenzzuweisung insgesamt 9 Units. Wiederum haben wir 10 Netzwerke für jeweils 100 Epochen trainiert, wobei eine Epoche einem Umfang von 10.000 Sätzen entspricht und das Trainings-Set nach jeweils zwei Epochen einmal komplett durchlaufen ist. Der Zugriff auf die Trainingssätze erfolgt innerhalb von zwei Epochen mit vollständiger Permutation. Nach jedem Satz werden die Kontext-Layer gelöscht.

<sup>43</sup>Für die Wortvorhersage-Komponente wurde die Größe des Hidden-Layers auf 80 Units und die Größe des Hidden-Layers der Referenzzuweisungs-Komponente auf 40 Units festgelegt. Für die Lernrate wurde  $\epsilon = 0.15$  gewählt, für das Momentum der Faktor *momentum* = 0.3, und für die initiale Randomisierung der Gewichte eine Grenze von *initialweightrange* = 0.1. Auch die Netzwerke in Simulation 2 erwiesen sich robust gegenüber einem breiten Spektrum weiterer Größenverhältnisse der Hidden-Layer der beiden Komponenten sowie innerhalb eines sinnvollen Rahmens anderer Trainingsparameter.

### 6.4.3. Ergebnisse

Die Darstellung der Ergebnisse erfolgt parallel zu Simulation 1. Zunächst erfolgt eine Analyse anhand der globalen Fehlermaße (GPE und RAE). Für die detaillierten Analysen der Output-Aktivierung ziehen wir diejenigen Positionen heran, die von besonderem Interesse sind, und bilden für die Auswertung der Referenzzuweisung entsprechende Referentenklassen. Abschließend werden wir wieder einen Blick auf die Strukturierung der beiden Hidden-Layer werfen.

#### 6.4.3.1. Performanz bei der Wortvorhersage

Für die Wortvorhersage-Komponente zeigt sich nach wenigen Epochen eine sehr gute Performanz. Nach etwa 30 Epochen zeigt sich keine nennenswerte Entwicklung des Fehlerwerts mehr. Das arithmetische Mittel über das gesamte Test-Set beläuft sich hier auf  $\bar{x}_{\text{GPEn}} = 0.103$  ( $\sigma = 0.077$ ). Der Vorhersagefehler für alle Verben beträgt  $\bar{x}_{\text{GPEn}} = 0.116$  ( $\sigma = 0.059$ ). Folglich wurden auch trotz der erhöhten Distanzen in Korpus 2 alle Fernabhängigkeiten gut gelernt. Zur detaillierten Analyse verwenden wir wieder die Entwicklung während der ersten 10 Epochen.<sup>44</sup>

**Komplexe Nebensätze** Abbildung 6.16 gibt den globalen Fehler (GPEn) gruppiert nach Epoche für vorangestellte Nebensätze mit eingebettetem Objekt-Relativsatz wieder. In Epoche 10 zeigt sich im Verlauf des Relativsatzes ein leicht erhöhter Fehlerwert für das Objekt und das Relativsatz-Verb (*koechin1 sieht. agr1*) und weiterhin für das Pronomen und das Verb im Matrixsatz. Zudem zeigt sich ein stark erhöhter Fehlerwert auf dem Objekt des folgenden Hauptsatzes. Für den Entwicklungsverlauf des Fehlerwerts auf verschiedenen Positionen ergeben sich folgende Auffälligkeiten: Der hohe Fehler auf dem Relativpronomen beschränkt sich auf die erste Epoche. Ein langfristiger Entwicklungsverlauf zeigt sich hingegen auf den drei Positionen nach dem Relativsatz (*sich kratzt. agr2 comma\_3*) sowie für das Subjekt des folgenden Hauptsatzes. Für die vorangestellten Nebensätze mit eingebettetem Subjekt-Relativsatz (und auch für die Kontrollstruktur mit Adverb) ergibt sich ein nahezu identischer Fehlerverlauf wie für den gezeigten Satztyp. Wir beschränken uns deshalb auf Abbildung 6.16.

Für die Fehler im Verlauf des Relativsatzes auf dem Relativpronomen und dem Subjekt bzw. Objekt lassen sich leider wiederum keine befriedigenden Antworten aus den Aktivierungen ableiten. Beim Relativpronomen verteilt sich die geringfügige falsche Aktivierung nach der ersten Epoche über mehrere Wortklassen (v.a. Verben und Artikel). Besonders beim Nomen müssen wir wieder von Ungleichmäßigkeiten der Aktivierungsverteilung innerhalb der aktivierten Nomen ausgehen (wenn ein korrektes Nomen zu stark und das andere zu schwach aktiviert ist, gehen beide Fehler in den Gesamtfehler ein). Der Fehler auf dem Relativsatzverb ist unter anderem auf die zu geringe Aktivierung des COMMA-Markers zurückzuführen, der auf eine zu geringe Erwartung eines weiteren eingebetteten Relativsatzes hinweist.

---

<sup>44</sup>Auch hier ist die Fehlerquote bereits für alle Wortpositionen (einschließlich der Verben) zufriedenstellend. Der Fehlerwert beträgt nach Epoche 10:  $\bar{x}_{\text{GPEn}} = 0.141$  ( $\sigma = 0.106$ ). Der Fehlerwert über alle Verben beträgt nach Epoche 10:  $\bar{x}_{\text{GPEn}} = 0.169$  ( $\sigma = 0.119$ ).

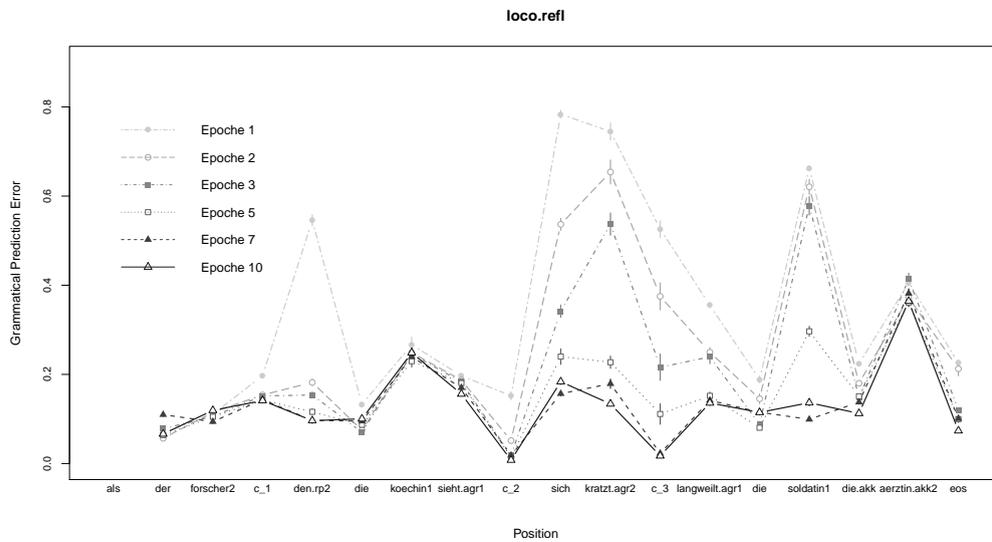


Abbildung 6.16.: Simulation 2. Verlauf und Entwicklung des globalen Fehlers der Wortvorhersage-Komponente, gruppiert nach Epoche: GPEn über Wortpositionen für vorangestellte Nebensätze mit zentral eingebettetem Objekt-Relativsatz (LSK-Bedingung).

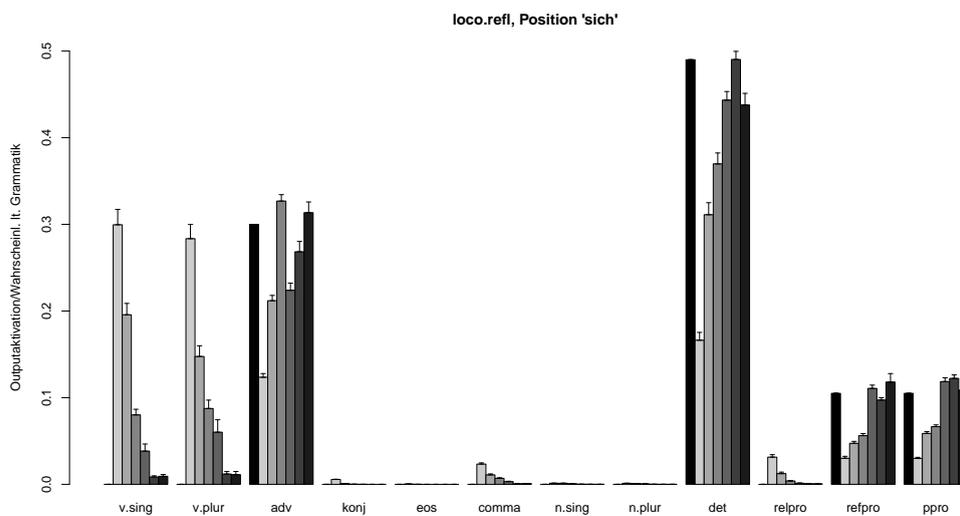


Abbildung 6.17.: Simulation 2. Details der Wort-Aktivierung im Vergleich zur absoluten Wortfolge-Wahrscheinlichkeit (Entwicklung über Epochen) für die Position der Anapher in vorangestellten Nebensätzen nach einem Objekt-Relativsatz (*Als der Forscher, den der Bäcker sieht, **sich** kratzt, ...*).

Von besonderem Interesse sind wiederum die Details zur Aktivierung für das finite Verb des Matrixsatzes nach dem eingebetteten Relativsatz, für das eine Fernabhängigkeit über 8 Positionen aufgelöst werden muss, sowie die Details zur Fehlerentwicklung auf der Anapher. Abbildung 6.17 zeigt die Aktivierungs-Details auf der Position der Anapher. Der hohe Fehler in den ersten Epochen beruht in erster Linie auf der Aktivierung der Verben. Er lässt sich darauf zurückführen, dass bei einer zentralen Einbettung in einen Hauptsatz ein Verb folgen müsste. In Nebensätzen muss aber entweder ein Adverb oder ein Objekt folgen, letzteres in Form eines Pronomens oder eines Artikels für ein Nomen im Akkusativ. Der anfänglich hohe Fehler ist also höchstwahrscheinlich ein Frequenz-Effekt der selteneren Nebensatzstruktur und einer schlechten Unterscheidung zwischen Haupt- und vorangestellten Nebensätzen (zumindest in dieser komplexen Form mit eingebettetem Relativsatz). Dieser Fehler zeigt sich identisch in den Kontrollätzen 1 und 2 und ist folglich unabhängig von der Form des eingebetteten Relativsatzes.

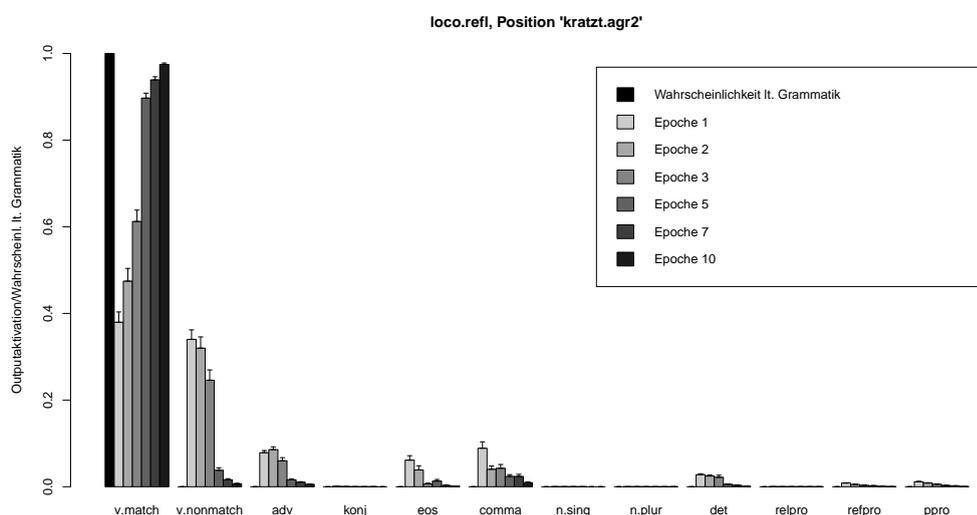


Abbildung 6.18.: Simulation 2. Details der Wort-Aktivierung im Vergleich zur absoluten Wortfolge-Wahrscheinlichkeit (Entwicklung über Epochen) für das Matrixsatz-Verb in vorangestellten Nebensätzen nach einem Objekt-Relativsatz (*Als der Forscher, den der Bäcker sieht, sich kratzt, ...*).

Die Details der Aktivierung für die Position des Matrix-Verbs zeigt Abbildung 6.18. In den frühen Epochen sind am Fehler Aktivierungen der falschen Wortklasse Adverb und Artikel sowie der Interpunktions-Marker EOS und COMMA beteiligt. In unserer Grammatik erscheinen Adverbien im Hauptsatz immer nach dem Verb und im Nebensatz nach dem Satz-Subjekt. Für Hauptsätze wäre nach einem Verb ein Adverb oder ein Artikel und nach einem Pronomen ein EOS oder COMMA-Marker möglich. Da für die vorausgegangene Position verstärkt auch Verben vorhergesagt wurden, weist dieser frühe Fehler wiederum auf die unzureichende Unterscheidung zwischen der selteneren Struktur vorangestellter Nebensätze und der Hauptsatzstruktur hin, die wir für die Position des Pronomens festgestellt haben. Größtenteils beruht der Fehler allerdings auf der Aktivierung der Verben mit falscher Kongruenz. Nach Epoche 5 ist diese Fehlaktivierung auf ein Minimum reduziert. Diese Leistung für die Auflösung der Subjekt-Prädikat-Kongruenz ist

bemerkenswert – immerhin muss für die hier gezeigte Position eine Fernabhängigkeit über sieben Positionen und über eine weitere davon unabhängige Subjekt-Prädikat-Kongruenz (Subjekt und Prädikat des Relativsatzes) überbrückt werden. Die exemplarisch gewählte Satzstruktur ist sehr komplex, und die erfolgreich aufgelöste Fernabhängigkeit belegt die erfolgreiche Verarbeitung der Subjekt-Prädikat-Kongruenz. Die Aktivierungsverteilungen sind für die anderen Satztypen mit vorangestelltem Nebensatz nahezu identisch. Der Fehler beruht also auf der generellen Komplexität und der nicht-kanonischen Wortfolge dieser Sätze.

Die auf den vorhergehenden Positionen vermutete Fehlleistung aufgrund einer selteneren irregulären Satzfolge der Satztypen mit vorangestelltem Nebensatz bestätigt sich auch im Fehler auf dem Komma (*comma\_3*) vor dem folgenden Hauptsatz: Dieser geht auf eine hohe Aktivierung des EOS-Markers zurück, der nach einem Hauptsatz oder einem regulären Nebensatz folgen könnte, nicht aber nach einem vorangestellten Nebensatz. Allerdings ist die Berücksichtigung der Voranstellung eine anspruchsvolle Aufgabe, da das Netzwerk für Korpus 2 zur korrekten Aktivierung über 13 Positionen hinweg die Information aufrechterhalten muss, dass der Nebensatz nicht auf einen Hauptsatz gefolgt ist. In Entsprechung zum Fehler für den seltenen vorangestellten Nebensatz findet sich auf dem nachfolgenden Hauptsatz ein sehr ähnlicher Fehler. Dieser Hauptsatz besitzt mit einem finiten Verb auf der ersten Satzposition eine sehr irreguläre Wortstellung. Zudem wird hier die Kongruenz-Klasse des Nomens durch das Verb bestimmt. Wie bereits in Simulation 1 gezeigt, führt diese Irregularität dazu, dass auf der Position des Satz-Subjekts (im Beispielsatz *soldatin1*) in den ersten Epochen gleichermaßen Nomen mit Prädikat-Kongruenz *agr1* und *agr2* aktiviert werden.

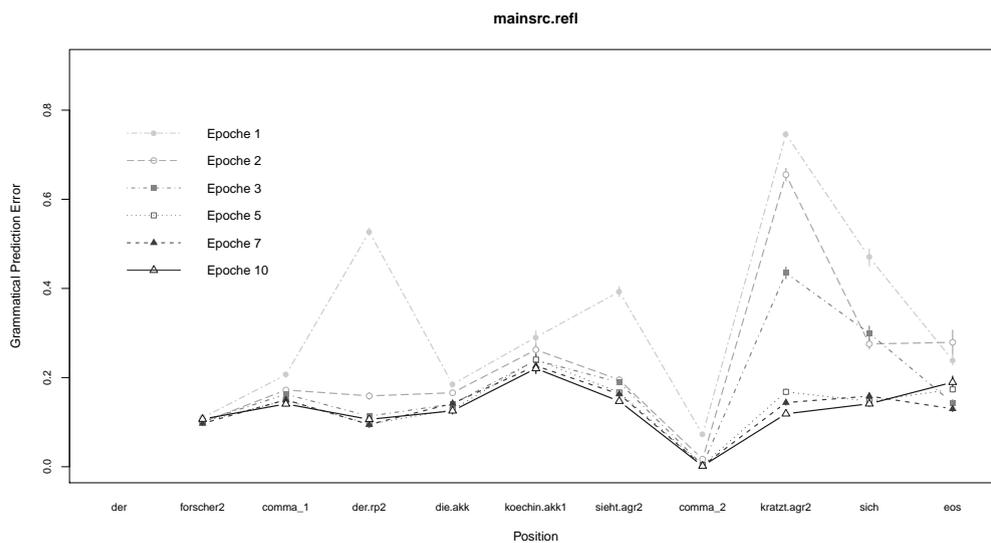


Abbildung 6.19.: Simulation 2. Verlauf und Entwicklung des globalen Fehlers der Wortvorhersage-Komponente, gruppiert nach Epoche: GPEn über Wortpositionen für Hauptsätze mit zentral eingebettetem Subjekt-Relativsatz (Kontrollbedingung 3).

**Komplexe Hauptsätze** Abbildung 6.19 zeigt den globalen Fehler (GPEn) gruppiert nach Epoche für Hauptsätze mit eingebettetem Subjekt-Relativsatz. Wie bereits bei den Nebensätzen kann kein Unterschied zwischen den Versionen mit Subjekt- gegenüber Objekt-Relativsatz festgestellt werden. Wir beschränken uns also auf den abgebildeten Satztyp. Wie bei den Nebensätzen zeigt sich im Verlauf des Relativsatzes ein auf Epoche 1 beschränkter Fehler auf dem Relativpronomen sowie bleibend leicht erhöhte Werte für das Nomen und das Relativsatz-Verb. Im Matrixsatz zeigt sich wie bereits beim Nebensatz eine langsame Verringerung des Fehlers auf den Positionen des Matrix-Verbs und des Pronomens, stabilisiert sich allerdings bereits ab Epoche 5 auf einem niedrigen Fehlerniveau.

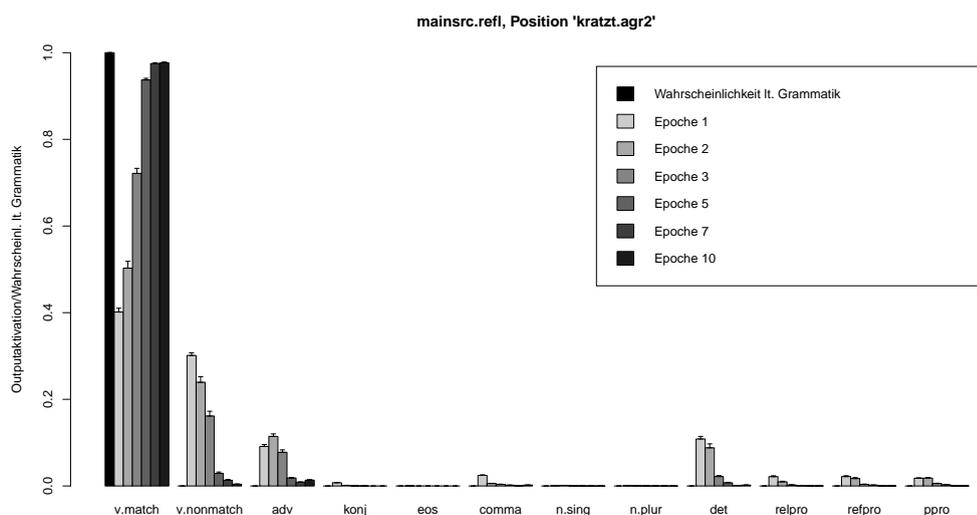


Abbildung 6.20.: Simulation 2. Details der Wort-Aktivierung im Vergleich zur absoluten Wortfolge-Wahrscheinlichkeit (Entwicklung über Epochen) für das Matrixsatz-Verb in Hauptsätzen nach einem Subjekt-Relativsatz (*Der Forscher, der den Bäcker sieht, kratzt sich.*).

Die Fehler auf den Elementen des Relativsatzes können auf dieselben Ursachen zurückgeführt werden, die bereits für die Versionen mit vorangestelltem Nebensatz beschrieben wurden. Es scheint keine Rolle zu spielen, in welchen Satztyp die Relativsätze eingebettet sind – Relativsätze werden demnach als syntaktisch eigenständige Sequenz verarbeitet. Interessanter sind die Aktivierungsverteilungen, die zum anfänglich erhöhten Fehler auf dem Matrixverb und dem Pronomen im Matrix-Satz führen. Abbildung 6.20 zeigt die Aktivierungsverteilung für das Matrix-Verb. Der anfängliche Fehler ist auf die Aktivierung der Adverbien und der Artikel sowie in geringerem Umfang der Pronomen zurückzuführen. Dies ist erstaunlich, denn diese Wortformen wären nur in einem Nebensatz möglich. Damit deutet sich an, dass sich die unzureichende Unterscheidung zwischen Haupt- und Nebensätzen in den frühen Epochen auch negativ auf die Wortvorhersage-Performanz in Hauptsätzen auswirkt. Abbildung 6.21 zeigt die Aktivierungsverteilung auf der Anapher. Hier beruht der Fehler der ersten drei Epochen vor allem auf der falschen Aktivierung des COMMA-Markers. Auch dies ist eine Vorhersage, die innerhalb eines Bigramms als Folgeelement eines Verbs nur nach einem Nebensatz korrekt wäre (wobei in diesem Fall

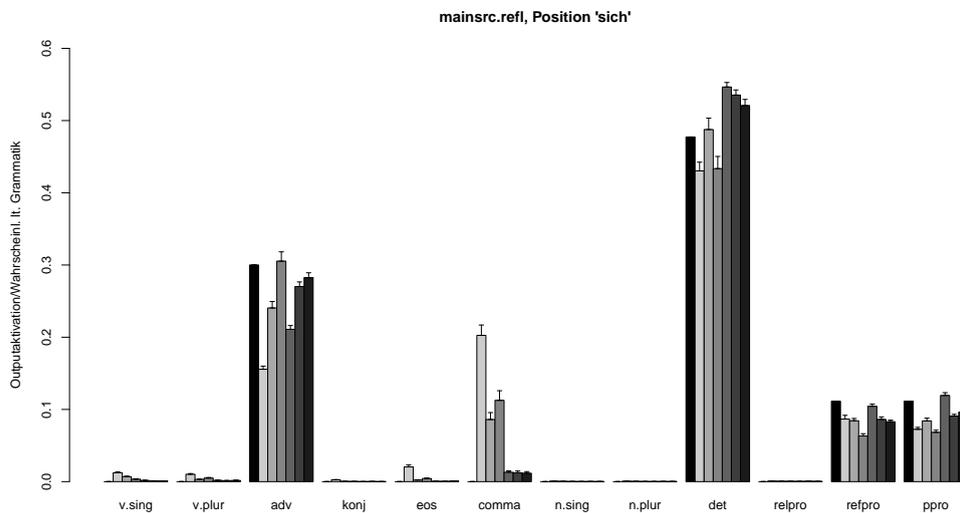


Abbildung 6.21.: Simulation 2. Details der Wort-Aktivierung im Vergleich zur absoluten Wortfolge-Wahrscheinlichkeit (Entwicklung über Epochen) für die Position der Anapher in Hauptsätzen nach einem Subjekt-Relativsatz (*Der Forscher, der den Bäcker sieht, kratzt sich.*).

auch eine stärkere Aktivierung des EOS-Markers zu erwarten wäre). Insgesamt scheint es sowohl für Haupt- als auch Nebensätze, dass zentral eingebettete Relativsätze die präzise Unterscheidung dieser Strukturen beeinträchtigt. Interessant ist schließlich noch der bleibend leicht erhöhte Fehler auf dem Komma (*comma\_1*), das den Relativsatz einleitet: In Simulation 1 haben wir diesen Fehler auf der Position des Relativpronomens gefunden, auf der die Verben übermäßig aktiviert wurden. Eine ähnliche Ursache hat der Fehler auf der Position des Kommas in Simulation 2. Da er hier aber über alle Epochen erhalten bleibt und sich zudem für andere Positionen mit Komma nicht ergibt, lässt er sich möglicherweise besser im Sinne eines Frequenz- oder Regularitäts-Effekts erklären.

#### 6.4.3.2. Performanz bei der Referenzzuweisung

Wie in Simulation 1 zeigt sich auch hier für die allgemeine Performanz bei der Referenzzuweisung bereits nach Epoche 10 ein sehr niedriger Fehler auf den meisten Satzpositionen, der aber einem hohen Fehlerwert für die pronominalen Formen gegenübersteht.<sup>45</sup> Die einzigen Wörter, auf denen die Referenzzuweisung kontextabhängig erfolgt, sind die Relativ-, Reflexiv- und Personalpronomen. Für diese zeigt sich auch nach einer hohen Anzahl von Epochen ein erhöhter Fehler. Erst nach ca. 90 Epochen zeigt sich keine substanzielle Verbesserung der Zuweisungsgenauigkeit mehr. Das arithmetische Mittel für die Summe der Pronominal-Positionen beläuft hier auf  $\bar{x}_{\text{RAEn}} = 0.011$  ( $\sigma = 0.053$ ), für alle pronominalen Formen auf  $\bar{x}_{\text{RAEn}} = 0.081$  ( $\sigma = 0.124$ ), und bei der ausschließlichen Berücksichtigung der Reflexiv- und Personalpronomen auf  $\bar{x}_{\text{RAEn}} = 0.155$  ( $\sigma = 0.134$ ).

<sup>45</sup>Fehlerwert nach Epoche 10:  $\bar{x}_{\text{RAEn}} = 0.033$  ( $\sigma = 0.14$ ). Fehlerwert für alle pronominalen Formen:  $\bar{x}_{\text{RAEn}} = 0.238$  ( $\sigma = 0.316$ ), ausschließlich für Reflexiv- und Personalpronomen:  $\bar{x}_{\text{RAEn}} = 0.448$  ( $\sigma = 0.314$ ).

**Globale Analyse** Für Korpus 2 lassen sich dieselben Merkmale feststellen wie bereits für Korpus 1: Die einzige Position, die Schwierigkeiten zu bereiten scheint und deren Fehler nicht gegen 0 tendiert, ist die Position der anaphorischen Referenz mit Personal- und Reflexivpronomen. Alle anderen Positionen einschließlich des Relativpronomens weisen ab Epoche 10 praktisch keine Fehlaktivierung mehr auf. Wir können uns also auch hier im weiteren Verlauf auf die Besprechung dieser Positionen beschränken. Auch der generelle Unterschied zwischen Personal- und Reflexivpronomen bei der Entwicklung des Fehlers zeigt sich wieder: Während Reflexivpronomen über einen langen Trainingsverlauf Probleme zu bereiten scheinen, ist der Fehler für Personalpronomen bereits in Epoche 10 mit einem Wert von ca. 0.3 relativ gering (vgl. Abb. 6.22 und 6.23). Allerdings zeigt sich für Reflexivpronomen langfristig eine Verbesserung in größeren Schritten, während die Verbesserungsleistung für Personalpronomen sehr gering ausfällt. Auf den letzten Epochen zeigt sich sogar ein geringerer Fehler für Reflexivpronomen (gemittelter Fehlerwert nach Epoche 90 bei unter 0.1) als für Personalpronomen (gemittelter Fehlerwert nach Epoche 90 knapp unter 0.2). Diese Entwicklung findet sich in den komplexen Satztypen unabhängig vom Matrixsatztyp und vom eingebetteten Relativsatztyp.

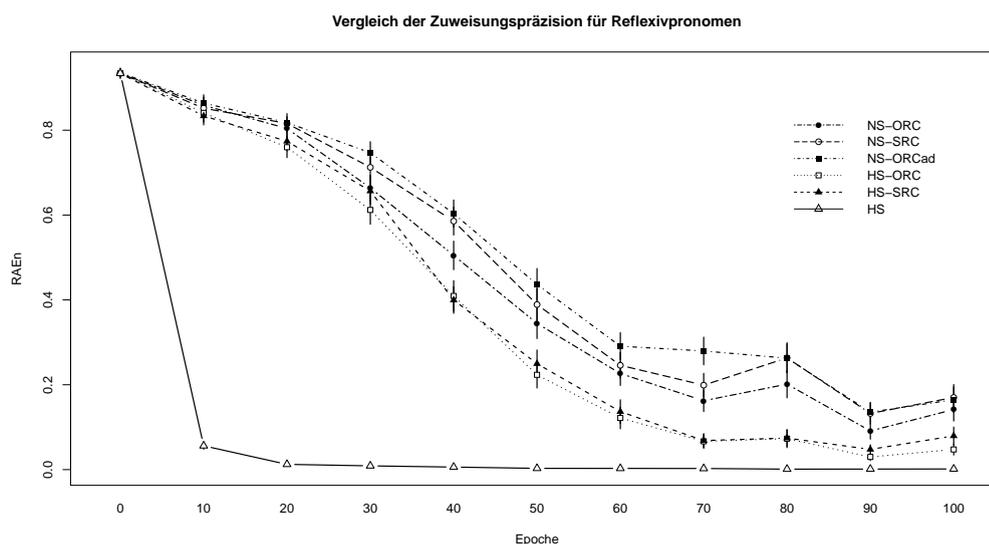


Abbildung 6.22.: Simulation 2. Vergleich der RAEn für Reflexivpronomen in verschiedenen Satztypen über den Verlauf der Epochen (Entwicklung).

Da in den Diagrammen über die verschiedenen Kontrollbedingungen mit bloßem Auge kein Unterschied in Entwicklung und Zuweisungsgenauigkeit auszumachen ist, stellen wir einen direkten Vergleich der Genauigkeit für die Anaphernposition über alle Satztypen dar. In den ersten beiden Diagrammen wird der globale Fehler für die Anaphernposition gruppiert nach Satztypen über alle Epochen in 10er-Schritten von der 10. bis zur 100. Epoche abgetragen, in Abbildung 6.22 für Reflexivpronomen, in Abbildung 6.23 für Perso-

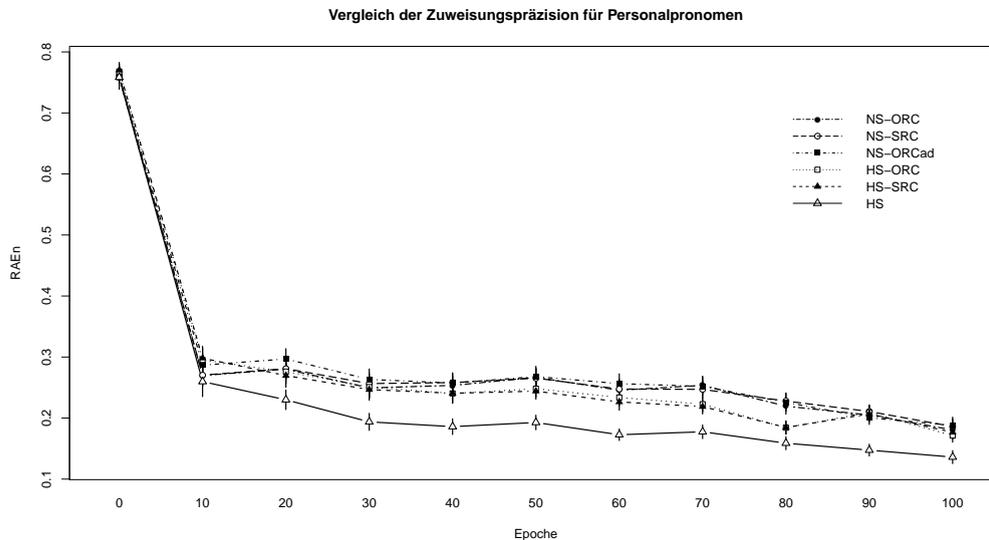


Abbildung 6.23.: Simulation 2. Vergleich der RAEn für Personalpronomen in verschiedenen Satztypen über den Verlauf der Epochen (Entwicklung).

nalpronomen.<sup>46</sup> Eine Rangliste der Satztypen nach der Qualität der Referenzzuweisung deutet sowohl für Reflexiv- als auch Personalpronomen folgende Abstufung an:

1. Die beste Performanz wird für Anaphern in einfachen Hauptsätzen erbracht. Für Reflexivpronomen zeigt sich dabei, dass die Leistung bereits nach 10 Epochen sehr gut ist, der globale Fehler beläuft sich auf einen Wert unter 0.1. Nach 20 Epochen zeigt sich keine Verbesserung mehr, der Fehler tendiert ab dieser Epoche gegen 0.0. Für Personalpronomen zeigt sich folgendes Bild: Während der Fehler nach Epoche 10 in allen Satzstrukturen einen ähnlichen Wert erreicht (knapp unter 0.3), trennt sich die Leistung in einfachen Hauptsätzen nach 20 Epochen von der Leistung in anderen Satzstrukturen und verbessert sich kontinuierlich bis Epoche 100 auf einen Wert von etwa 0.13. Die Leistung ist dabei über alle Epochen hinweg besser als für die übrigen Satzstrukturen.
2. Für Anaphern in Hauptsätzen nach eingebettetem Relativsatz zeigt sich eine etwas schlechtere Performanz. Dabei kann kein wesentlicher Unterschied zwischen der Art des eingebetteten Relativsatzes (Subjekt- oder Objekt-Relativsatz) festgestellt werden. Für Reflexivpronomen zeigt sich dabei im Unterschied zu einfachen Hauptsätzen zu Beginn ein hoher Fehler (über 0.8), der kontinuierlich abnimmt und seinen geringsten Wert nach Epoche 90 mit einem gemittelten RAEn deutlich unter 0.1 erreicht. Die Entwicklung für Personalpronomen verläuft dagegen ähnlich zur Entwicklung in einfachen Hauptsätzen. Der Fehler sinkt von einem Fehlerwert knapp

<sup>46</sup>Zur kurzen Erläuterung der Legende: Das Kürzel *NS-ORC* steht für den Satzkontext mit Objekt-Relativsatz, der in einen Nebensatz eingebettet ist, *NS-SRC* und *NS-ORCad* stehen für die ersten beiden Kontrollbedingungen (Nebensatz mit Subjekt-Relativsatz oder Objekt-Relativsatz erweitert um ein Adverb), *HS-ORC* und *NS-SRC* stehen für die weiteren Kontrollbedingungen 3 und 4 (Einbettung eines Objekt- oder Subjekt-Relativsatzes in einen Hauptsatz), *HS* schließlich steht für einen einfachen Hauptsatz.

unter 0.3 nach Epoche 10 auf einen Fehlerwert von knapp unter 0.2 nach Epoche 100.

3. Die schlechteste Performanz wird in Nebensätzen mit eingebetteten Relativsätzen erreicht. Hier findet sich zudem ein geringfügiger Unterschied des Einflusses des intervenierenden Materials vor der Anapher. Vor allem die Version mit eingebettetem Objekt-Relativsatz und zusätzlichem Adverb schneidet während des Verlaufs minimal schlechter ab. Auf den Epochen 90 und 100 aber hat sich der Fehler für alle drei Versionen angeglichen und erreicht sein Minimum auf Epoche 90 mit einem Wert von etwa 0.15. Insgesamt gleicht der Verlauf der Entwicklung dem Verlauf bei Hauptsätzen mit Relativsatz. Auch für die Personalpronomen verläuft die Entwicklung ähnlich wie im Kontext von Hauptsätzen mit eingebettetem Relativsatz. Sie nimmt den gleichen Verlauf mit stellenweise geringfügig höherem Fehler und endet auf den Epochen 90 und 100 auf demselben Fehlerwert wie diese (knapp unter 0.2).

Abbildung 6.24 vergleicht das arithmetische Mittel des RAEn der letzten fünf Epochenstufen (Epoche 60-100)<sup>47</sup> für die Anapherntypen gruppiert nach Satztyp. Insgesamt scheint der Fehler vor allem von der Klasse des Satztyps des Matrixsatzes und vom Vorhandensein intervenierenden Materials abzuhängen. Dies zeigt sich deutlicher für die Reflexivpronomen. Bei den Personalpronomen setzt sich in erster Linie der einfache Hauptsatz ab. Die Art des intervenierenden Materials scheint einen vergleichsweise geringen Einfluss auf die Auflösung der anaphorischen Referenz zu nehmen. Der etwas höhere Fehler für die Objekt-Relativsatz-Version mit zusätzlichem Adverb könnte mit der zusätzlichen Distanz zusammenhängen, die hier zur Auflösung der Fernabhängigkeit überbrückt werden muss. Erstaunlich ist aber, dass Hauptsätze mit eingebettetem Relativsatz besser abschneiden als entsprechende Nebensätze, denn erstere erzwingen die Überbrückung einer geringfügig längeren Fernabhängigkeit (8 vs. 7 Positionen).

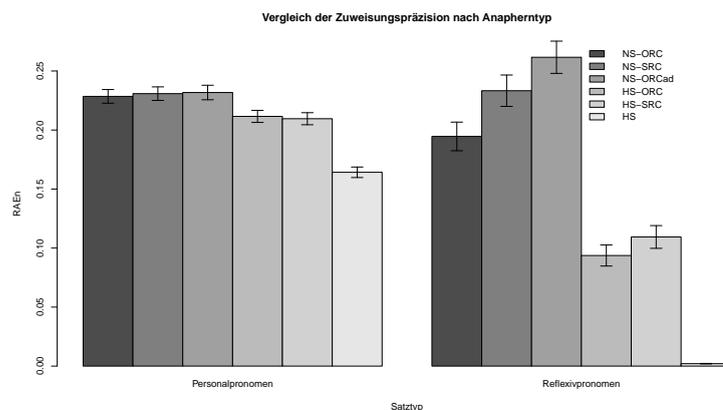


Abbildung 6.24.: Simulation 2. Vergleich des RAEn für verschiedene anaphorische Formen (Reflexivpronomen und Personalpronomen), gemittelt über die Epochen 60-100, gruppiert nach Satztyp (sortiert von komplex zu einfach).

<sup>47</sup>Diese Epochen wurden ausgewählt, da in diesem Spektrum der globale Fehler für Reflexivpronomen hinreichend stabil und damit eine bessere Vergleichbarkeit zu den Personalpronomen gegeben ist.

**Detailanalyse: Reflexivpronomen** Zur detaillierten Untersuchung der Aktivierung für die Referenzzuweisung haben wir wiederum Referentenklassen erstellt, über die wir in den Output-Aktivierungen mögliche unterschiedliche Ursachen für den Fehler der Anapherauflösung identifizieren können. Die Referentenklassen fallen etwas anders aus als in Simulation 1, da wir mit der genus-unabhängigen Subjekt-Prädikat-Kongruenz und der unterschiedlichen Markierung der intervenierenden Nomen und Verben weitere Merkmale berücksichtigen müssen, die die Auflösung beeinflussen könnten. Wir erläutern die Referentenklassen wieder an einem Beispielsatz: *Als der Forscher, den der Minister sieht, sich kratzt, langweilt die Soldatin die Ärztin.* Für die Reflexivpronomen unterscheiden wir die folgenden Klassen:

- Antezedens (1\_Ante): Subjekt des Matrixsatzes und damit korrektes Antezedens des Reflexivpronomens (*Forscher*).
- Distraktor (2\_Distr): Referent des im Relativsatz genannten Eigennamens und damit direkt vor dem Pronomen genanntes Subjekt oder Objekt (*Minister*, oder jeder andere Referent, der auf dieser Position erscheint – außer *Forscher*).
- Distraktor mit Verb- und Genus-Kongruenz (3\_DistrK): Nochmals der Referent des im Relativsatz genannten Eigennamens. Allerdings werden in dieser Klasse nur diejenigen Referenten berücksichtigt, die in Genus und Prädikat-Kongruenz mit den Merkmalen des jeweils korrekten Antezedens übereinstimmen (nur *Minister* ist im Beispielsatz möglich, da er in Genus und Prädikat-Kongruenz mit *Forscher* übereinstimmt).
- Referenten mit gleicher Verb- und Genus-Kongruenz (4\_VerbGenK): Mittelwert der Aktivierung aller Referenten, die in Genus und Prädikat-Kongruenz mit dem korrekten Antezedens übereinstimmen. Es werden alle Referenten mit entsprechenden Eigenschaften berücksichtigt außer dem Antezedens selbst (wiederum *Minister*, auch dann, wenn er nicht wie im Beispiel im Relativsatz erscheinen würde).
- Referenten mit gleicher Verb-Kongruenz (5\_VerbK): Mittelwert der Aktivierung aller Referenten, deren Prädikat-Kongruenz mit der des korrekten Antezedens übereinstimmt, wobei das Antezedens selbst nicht berücksichtigt wird (*Minister*, *Ärztin* und *Richterin*).
- Referenten mit gleicher Genus-Kongruenz (6\_GenusK): Mittelwert der Aktivierung aller Referenten, deren Genus mit dem des korrekten Antezedens übereinstimmt (das Antezedens selbst wird nicht berücksichtigt) (*Minister*, *Bäcker* und *Handwerker*).
- Nichtkongruente Referenten (7\_NichtK): Mittelwert der Aktivierung aller Referenten, deren Genus nicht mit dem des korrekten Antezedens übereinstimmt (*Soldatin*, *Köchin*, *Ärztin* und *Richterin*).

Die Abbildungen 6.25 und 6.26 zeigen die Aktivierungen für die Referentenklassen über die Epochen 10 bis 80. Wir legen die im paarweise vergleichenden t-Test (mit Korrektur nach Holm) signifikanten Unterschiede (mindestens  $p < 0.05$ ) dar. Der Unterschied für die korrekte Aktivierung des Antezedens (*Forscher*) wird zwischen den einfachen Hauptsätzen und den komplexen Sätzen bis Epoche 50 signifikant. Ab Epoche 60 verliert sich der Unterschied zwischen den verschiedenen Hauptsatztypen. Der Unterschied zwischen einfachen Hauptsätzen und den Nebensatztypen bleibt hingegen signifikant bis einschließlich Epoche

80. Zwischen den komplexen Satztypen zeigen sich erste Unterschiede ab Epoche 30. In Epoche 40 werden die Unterschiede zwischen den Kontrollsätzen 1 und 2 gegenüber den Hauptsätzen mit Relativsatz signifikant (nicht aber der Unterschied zwischen diesen Hauptsätzen und der LSK-Bedingung NS-ORC). Dieser Unterschied bleibt ebenfalls bis einschließlich Epoche 80 erhalten (außer in Epoche 60 zwischen NS-SRCK und HS-SRC). Zudem wird in Epoche 70 das Antezedens in der LSK-Bedingung signifikant besser erkannt als in Kontrollbedingung 2. Insgesamt zeichnet sich hier ab, dass die anaphorische Verarbeitung in den komplexen Nebensätzen am schlechtesten erfolgt. Gerade für die LSK-Bedingung zeigt sich dieses Verhältnis aber nicht, was auf einen eher kontextunabhängigen Effekt hinweist. Dies ist unerwartet, da der Satztyp NS-ORC zumindest die gleiche Komplexität aufweisen sollte wie die Kontrollsätze 1 und 2 (NS-SRCK und NS-ORCK).

Für die folgende Diskussion ziehen wir immer den Vergleich zwischen den Referentenklassen unabhängig vom Satz heran, um die relevanten Klassen zu identifizieren. Zugleich ziehen wir den Vergleich innerhalb der Klassen heran, um Unterschiede zwischen den Satztypen zu identifizieren. Zwischen Distraktoren und kongruenten Distraktoren finden wir das gleiche Verhältnis wie in Simulation 1: Über alle Satztypen hinweg erhält die Klasse der allgemeinen Distraktoren in den ersten Epochen mehr Aktivierung als die zum Antezedens kongruenten Distraktoren. Der kongruente Distraktor (*Minister*) scheint also besser als strukturell auszuschließender Antezedens erkannt zu werden und erhält dementsprechend weniger Aktivierung. Zwischen den komplexen Satztypen werden innerhalb der beiden Distraktorklassen allerdings keine Unterschiede signifikant.

Bei den allgemeinen Referenten – d.h. für die Aktivierung der Referenten unabhängig von ihrem Erscheinen im Satz (außer dem Antezedens) – zeigen sich über alle Satztypen hinweg ebenfalls Unterschiede: Von Epoche 10 bis 60 werden die Referenten mit vollständig übereinstimmender Kongruenz (wiederum *Minister*, aber auch dann, wenn er nicht im Beispielsatz erscheinen würde) signifikant stärker aktiviert als die Referenten, die nur hinsichtlich der Prädikat-Kongruenz mit dem Antezedens übereinstimmen (5\_VerbK). Diese wiederum werden stärker aktiviert als die Referenten, die nur hinsichtlich des Genus mit dem Antezedens übereinstimmen (6\_GenusK). Folglich ist Prädikat-Kongruenz für das Netzwerk eine wesentlich wichtigere oder präsentere Information als das Genus. Der Unterschied der Klasse der Referenten mit vollständig übereinstimmenden Kongruenzmerkmalen (4\_VerbGenK) gegenüber den anderen beiden Klassen bleibt auch in den späteren Epochen erhalten. Wir gehen davon aus, dass die Unterschiede zwischen den verschiedenen Satztypen in den weniger beschränkten Klassen nur Unterschiede der Klasse 4\_VerbGenK reflektieren. Da diese Klasse zudem bis Epoche 70 stärker aktiviert wird als die der Distraktoren, ist davon auszugehen, dass die intervenierenden Informationen keine Rolle für die Anaphernauflösung spielen. Vielmehr scheinen sie als strukturell unzugängliche Antezedenzen erkannt zu werden. Die Interferenz-Effekte hingegen sind reine Effekte des Abrufs der relevanten Informationen in Interaktion mit der unterschiedlichen Komplexität der verschiedenen Satztypen. Wir reduzieren die nächsten Vergleiche deshalb auf die Klasse der Referenten mit vollständiger Kongruenzübereinstimmung: Der signifikante Unterschied der einfachen Hauptsätze zu den komplexen Sätzen bleibt bis Epoche 40 erhalten. Ab Epoche 50 wird die falsche Aktivierung der vollständig kongruenten Referenten nur noch für die komplexen Nebensatz-Strukturen (außer der LSK-Bedingung) signifikant. Ab Epoche 60 verliert sich auch dieser Unterschied (außer für NS-ORCK in Epoche 70

und NS-SRCK in Epoche 80). Für die komplexen Satztypen zeigen sich für die Epochen 40 und 50 Unterschiede, die weitgehend umgekehrt parallel zu den Unterschieden der Aktivierung des Antezedens verlaufen: In den Kontrollsätzen 1 und 2 (NS-SRCK und NS-ORCK) werden die Referenten stärker aktiviert als in den komplexen Hauptsatz-Strukturen. Allerdings findet sich dieser Unterschied nur noch bedingt in den späteren Epochen (in Epoche 70 zwischen NS-ORCK und HS-ORC, in Epoche 80 zwischen NS-SRCK und HS-SRC/HS-ORC).

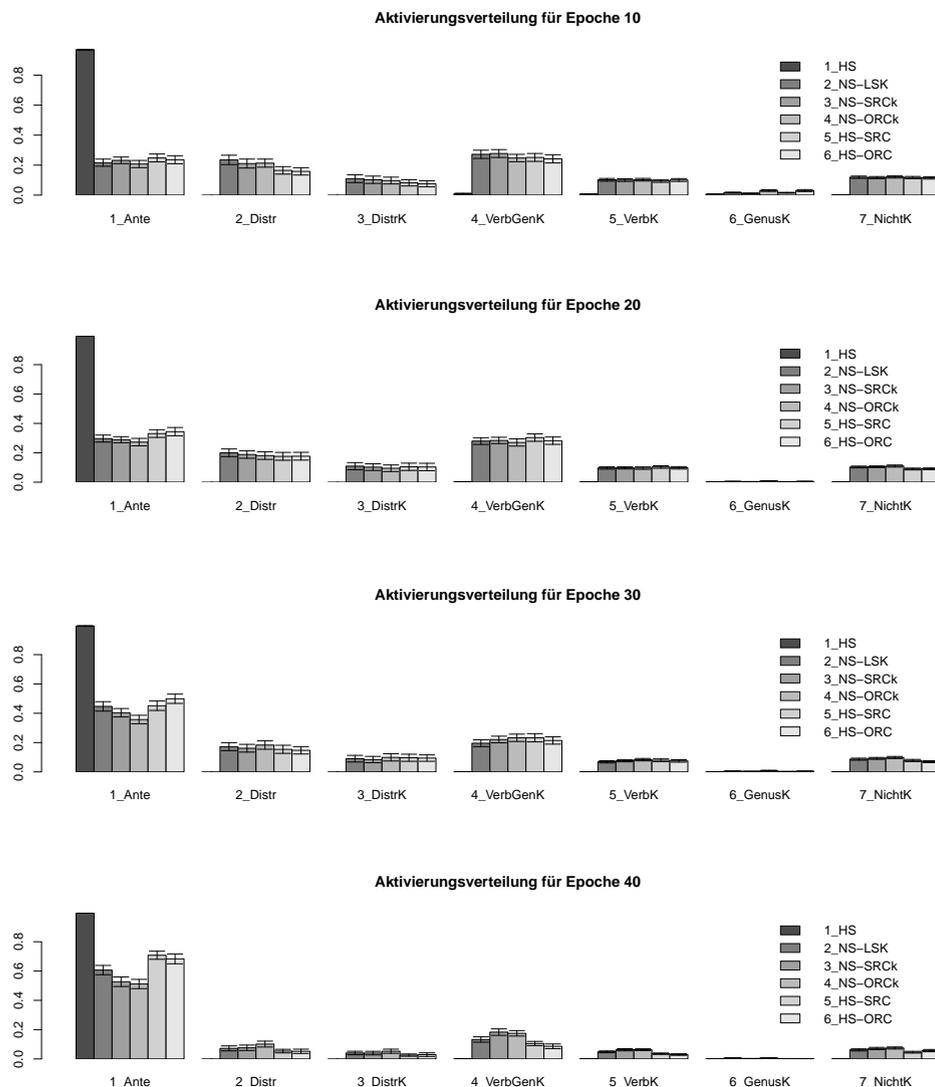


Abbildung 6.25.: Simulation 2. Details der Referenten-Aktivierung für Reflexivpronomen, Epochen 10 bis 40.

**Detailanalyse: Personalpronomen** Die Referentenklassen des Personalpronomens entsprechen denen, die wir in Simulation 1 eingeführt haben. Wir erläutern sie nochmals anhand eines angepassten Beispielsatzes:

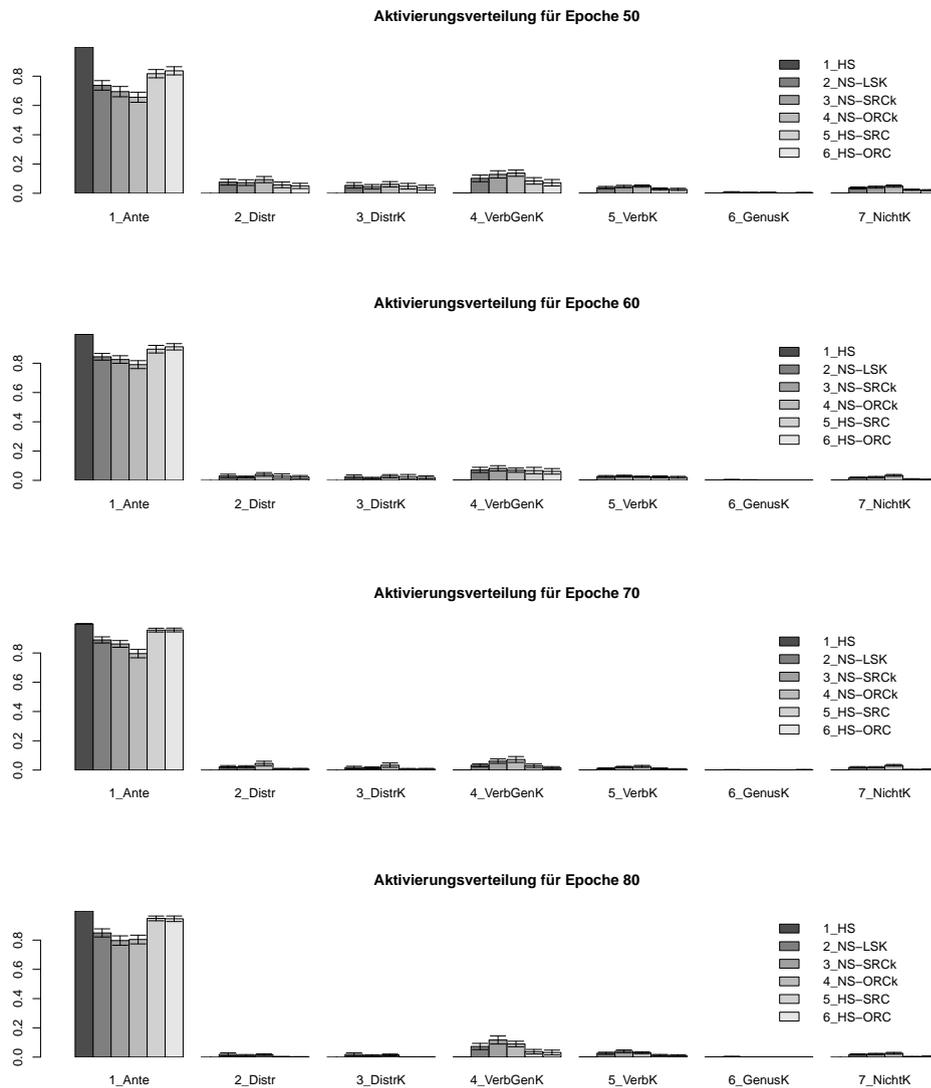


Abbildung 6.26.: Simulation 2. Details der Referenten-Aktivierung für Reflexivpronomen, Epochen 50 bis 80.

*Als der Forscher, den der Minister sieht, ihn kratzt, langweilt die Soldatin die Ärztin.*

- Antezedenzien (1\_Antezedens): Mittelwert aller nach syntaktischen und morpho-syntaktischen Bedingungen mögliche Antezedenzien des Personalpronomens (alle Referenten, deren Genus mit dem des Pronomens übereinstimmen, also *Minister Bäcker Handwerker*, nicht aber das Subjekt des Matrixsatzes *Forscher*).
- Distraktor (2\_Distraktor): Referent des im Relativsatz genannten Eigennamens und bei Genus-Kongruenz zum Pronomen möglicher Antezedens (*Minister*).
- Subjekt des Matrixsatzes (3\_Matrixsubjekt): Subjekt des Matrixsatzes, der das Pronomen enthält, und damit als Antezedens zwingend ausgeschlossen (*Forscher*).

Abbildung 6.27 zeigt die Aktivierungen der Referentenklassen des Personalpronomens für die Epoche 10 bis 80. Beginnen wir mit dem Antezedens: Der Unterschied zwischen einfachen Hauptsätzen und komplexen Satztypen wird ab Epoche 20 signifikant und bleibt für die komplexen Hauptsatz-Strukturen bis Epoche 60 und für die komplexen Nebensatz-Strukturen bis Epoche 80 erhalten. Zwischen den verschiedenen komplexen Satzstrukturen können hingegen keine Unterschiede festgestellt werden. Dieses Bild wiederholt sich umgekehrt bei der falschen Aktivierung des Matrixsatz-Subjekts: Ab Epoche 30 bis Epoche 80 wird dieses in einfachen Hauptsätzen signifikant weniger stark aktiviert als in allen komplexen Satztypen. Zwischen den verschiedenen komplexen Satzstrukturen finden sich wiederum keine signifikanten Unterschiede, ebenso nicht für die Klasse der Distraktoren. Wie bei den Reflexivpronomen ist wiederum der Unterschied zwischen den Referentenklassen gemittelt über die verschiedenen Satztypen interessant: Zum einen ist die Aktivierung der korrekten Antezedenzien in allen Epochen signifikant stärker als die des Matrixsatz-Subjekts. Die koreferenzielle Auflösung erfolgt also bereits ab Epoche 10 korrekt. Zum anderen ist die Aktivierung der korrekten Antezedenzien in den Epochen 50 bis 70 (gemittelt über Satztypen, ohne einfache Hauptsätze) signifikant größer als die der ebenfalls korrekten Distraktoren. Mit aller Vorsicht kann das als Hinweis gedeutet werden, dass der Distraktor wie in Simulation 1 im Sinne eines potenziellen Satz-Subjekts einer LSK tendenziell geringer aktiviert wird.

### 6.4.3.3. Interne Repräsentation

Werfen wir einen Blick auf die interne Strukturierung für das komplexere Korpus 2: Die folgenden Diagramme veranschaulichen die Struktur der internen Repräsentationen auf Grundlage der Aktivierungsmuster der Hidden-Layer der beiden Komponenten. Die ersten beiden Diagramme (Abb. 6.28 und 6.29) geben die Struktur des Hidden-Layers der Wortvorhersage-Komponente über die Mittelwert-Vektoren der Wörter und die anhand des Satztyps spezifizierten Pronomen wieder. Die beiden darauffolgenden Diagramme (Abb. 6.30 und 6.31) zeigen die entsprechenden Strukturen für das Hidden-Layer der Referenzzuweisungs-Komponente. Als Grundlage verwenden wir wieder das Test-Korpus.<sup>48</sup> Wie zuvor für Simulation 1 haben wir auch für Simulation 2 verschiedene strukturentdeckende Verfahren eingesetzt. Wir beschränken uns aber auf die Abbildung der jeweiligen HCA, die zur Erfassung der relevanten Strukturen genügt. In den Abbildungen

<sup>48</sup>Aus diesem Grund fehlen in den Diagrammen für die Wortformen die Konjunktionen *während* und *und* sowie das Verb *zeichnet* in seinen beiden Konjugationsformen.

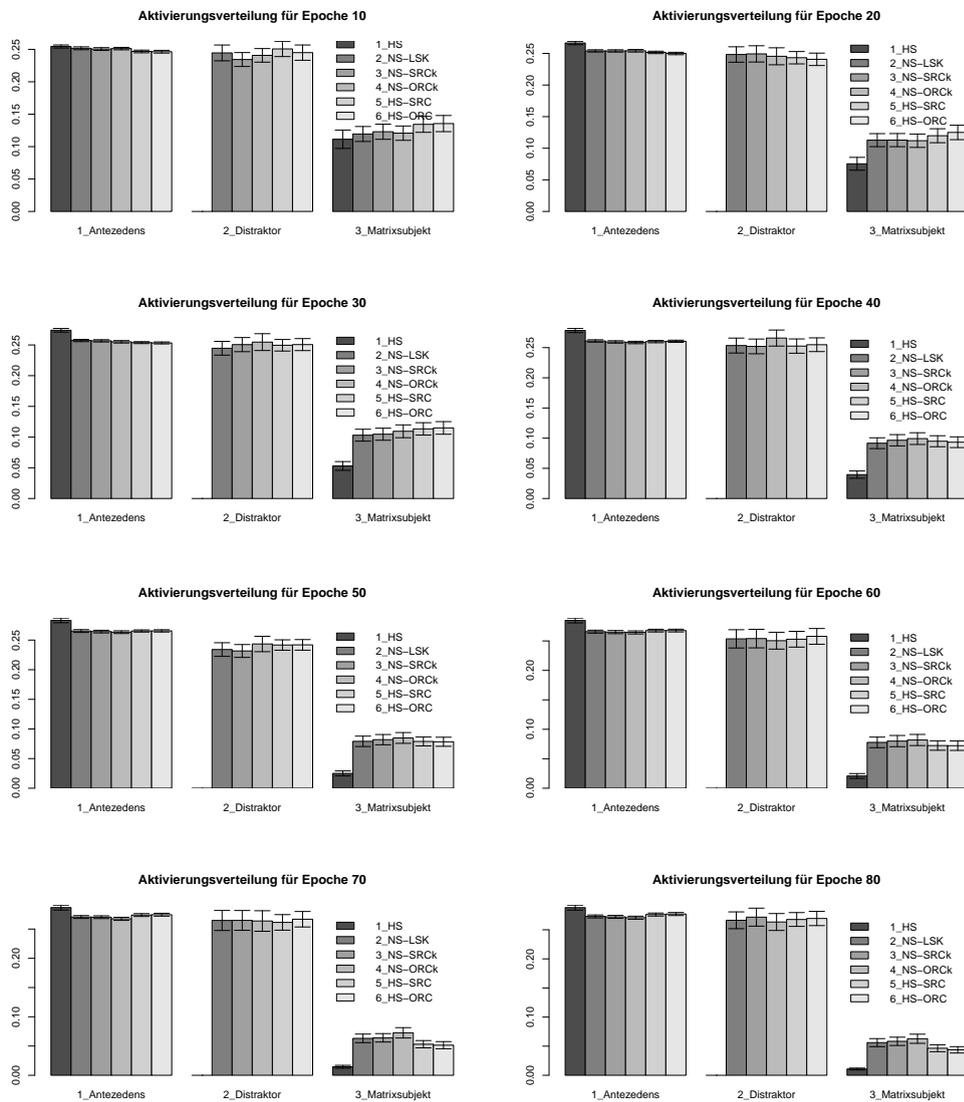


Abbildung 6.27.: Simulation 2. Details der Referenten-Aktivierung für Personalpronomen.

6.29 und 6.30 haben wir wieder Kurzbezeichnungen der verschiedenen syntaktischen Kontexte verwendet. Diese sind in nachstehender Liste aufgeführt:

- Einfacher Hauptsatz: HC\_
- Hauptsatz mit eingebettetem Subjekt-Relativsatz: SRC\_ (Kontrollbedingung 3)
- Hauptsatz mit eingebettetem Objekt-Relativsatz: ORC\_ (Kontrollbedingung 4)
- Nebensatz mit eingebettetem Objekt-Relativsatz: ORClc\_ (LSK-Bedingung)
- Nebensatz mit eingebettetem Subjekt-Relativsatz: SRCgl\_ (Kontrollbedingung 1)
- Nebensatz mit eingebettetem Objekt-Relativsatz mit Adverb: ORCad\_ (Kontrollbedingung 2)

**Wortvorhersage-Komponente** Die Abbildung 6.28 stellt die Clusterung der Wörter anhand einer agglomerativen HCA dar, Abbildung 6.29 zeigt die Struktur für die satztypabhängige Aktivierung durch die Pronomen. Wiederum zeigt sich für die Mittelwert-Vektoren die Kategorisierung nach Wortarten: Die Nomen sind gemeinsam gegliedert, dabei findet sich als erste Feingliederung die Unterscheidung des Kasus nach Akkusativ und Nominativ. Nomen im Nominativ unterscheiden sich zudem nach ihrer Prädikat-Kongruenz-Klasse. Die Relativpronomen werden ebenfalls zusammen gruppiert, unterscheiden sich aber zuerst nach ihrer Prädikat-Kongruenz-Klasse und erst auf der untersten Stufe nach ihrem Kasus. Insgesamt scheint bei diesen Clustern das Genus keine Rolle zu spielen. Dies ist insofern nachvollziehbar, als dass die Subjekt-Prädikat-Kongruenz (in Korpus 2 im Unterschied zu Korpus 1) nicht durch dieses Merkmal bestimmt wird. Als nächstes folgt ein Cluster für die Verben. Der letzte Cluster unterscheidet schließlich in seiner Feingliederung zwischen Adverbien, Pronomen, Artikeln sowie Interpunktionszeichen und Konjunktionen.

Abbildung 6.29 zeigt die aus Simulation 1 bekannte Clusterung der Pronomen in Abhängigkeit ihres syntaktischen Kontexts. In dieser unterscheiden sich zunächst die Haupt- und Nebensätze, erst auf den Folgestufen werden zuerst die verschiedenen Pronomen und schließlich die Details der Satzstruktur differenziert. Dass gerade diese Detailunterscheidung die geringste Differenzierung im Zustandsraum erfährt, spiegelt das Maß der Generalisierung unabhängig von linearen Ähnlichkeiten wider.

**Referenzzuweisungs-Komponente** In Abbildung 6.30 finden wir wiederum die genau umgekehrte Strukturierung in der Referenzzuweisungs-Komponente: In dieser wird zuerst zwischen den Pronomen unterschieden (eine eFA zeigt, dass die Reflexivpronomen im Zustandsraum eher zwischen den beiden Genus-Formen verortet sind). Erst die nächste Clusterebene unterscheidet zwischen Haupt- und Nebensätzen.

Für die HCA des Hidden-Layers der Referenzzuweisungs-Komponente haben wir die Lösung des divisiven Cluster-Verfahrens verwendet, da dieses besser die strukturellen Merkmale wiedergibt, die wir auch bei einer eFA und einer PCA beobachten konnten. Abbildung 6.31 zeigt die Cluster-Lösung für die Mittelwert-Vektoren der unterscheidbaren Wörter. Die Strukturierung unterscheidet wieder zunächst zwischen den für die Referenzzuweisung relevanten und irrelevanten Wörtern. In der Gruppierung der referenzrelevanten Wörter wird zunächst anhand des Genus zwischen femininen und maskulinen Wortfor-

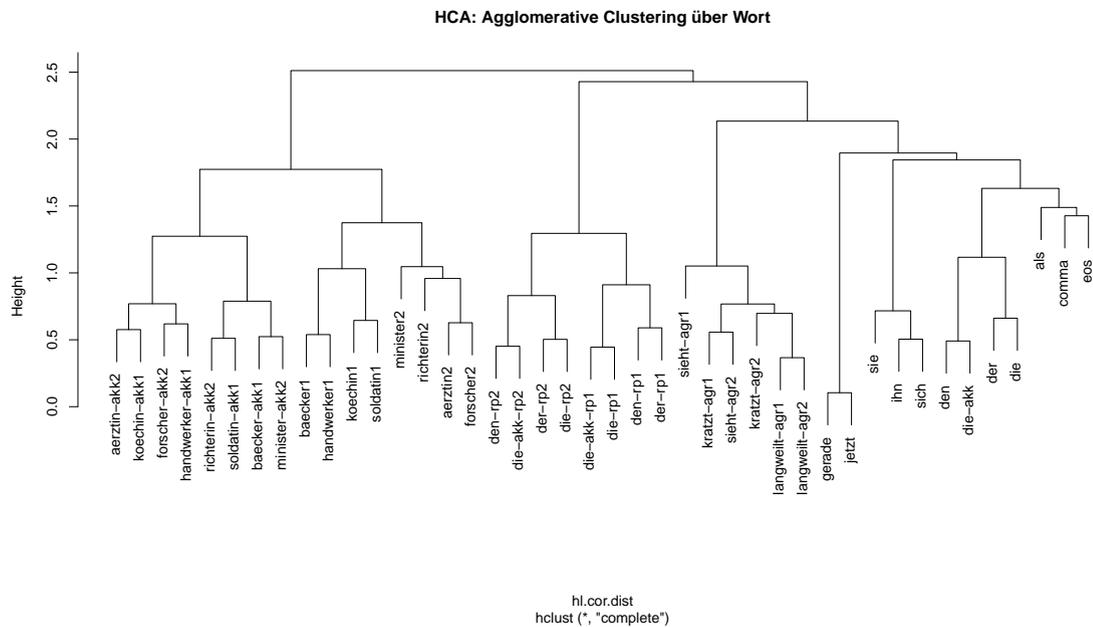


Abbildung 6.28.: Simulation 2. Hierarchische Cluster-Analyse (agglomerativ) der Hidden-Layer-Aktivierungen der Wortvorhersage-Komponente: Räumliche Gliederung der Wortformen.

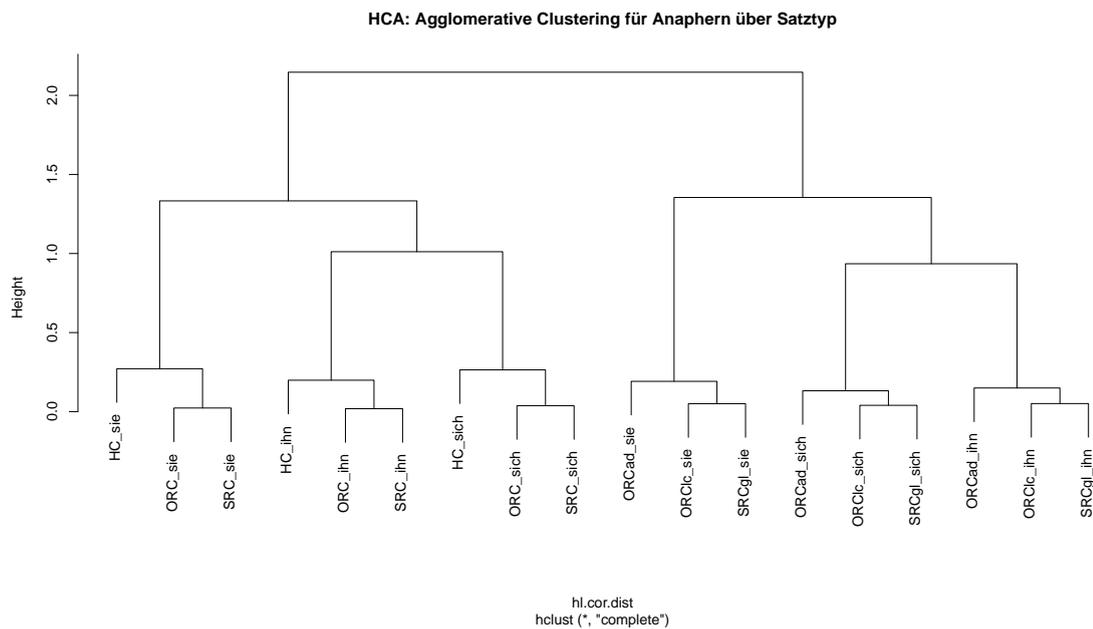


Abbildung 6.29.: Simulation 2. Hierarchische Cluster-Analyse (agglomerativ) der Hidden-Layer-Aktivierungen der Wortvorhersage-Komponente: Anaphern in Abhängigkeit ihres syntaktischen Kontexts.

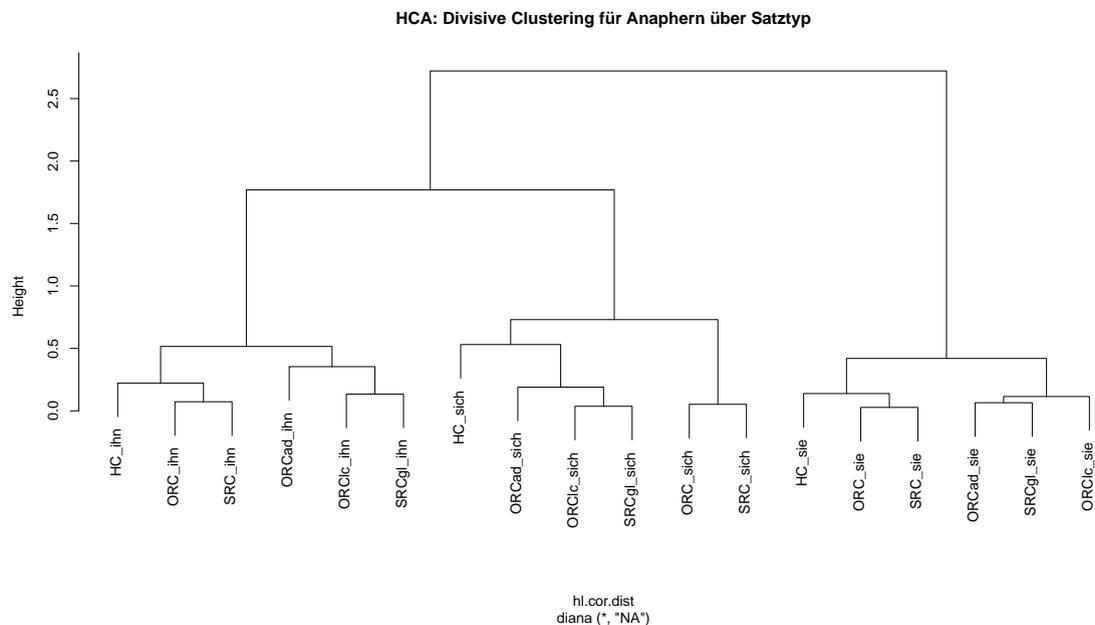


Abbildung 6.30.: Simulation 2. Hierarchische Cluster-Analyse (divisives Clusterbildungsverfahren) der Hidden-Layer-Aktivierungen der Referenzzuweisungs-Komponente: Anaphern in Abhängigkeit ihres syntaktischen Kontexts.

men unterschieden, in der Feingliederung schließlich zeigt sich, dass die Wortformen einzelner Referenten unabhängig vom Kasus zusammen gruppiert werden. In einer eFA (ohne Abb.) zeigt sich wiederum, dass die Zuordnung des Reflexivpronomens eher in einer Zwischenstellung zwischen den Genera erfolgt.

#### 6.4.4. Diskussion

Insgesamt zeigt sich wieder eine sehr gute allgemeine Performanz bei der Wortvorhersage. Die im Vergleich zu Simulation 1 wesentlich längeren Distanzen durch Artikel und Kommata sowie die stark erweiterte grammatische Markierung durch den Kasus für Artikel und Nomen sowie die Prädikat-Kongruenz für Nomen, Relativpronomen und Verben scheinen keinerlei Probleme zu bereiten. In den Aktivierungs-Details der Wortvorhersage finden wir wiederum vor allem im Verlauf der Relativsätze und in der folgenden Region des Matrixsatzes erhöhte Fehlerwerte. Sie gehen in den verschiedenen Fällen wieder auf die in Kapitel 6.3.4 besprochenen Effekte zurück. Allerdings finden sich praktisch keine Fehler mehr, die im Sinne eines LSK-Effekts erklärt werden könnten. Die Satzstruktur-Informationen durch die Kommata scheinen also ihren Zweck in dieser Hinsicht zu erfüllen.

Auch für die allgemeine Referenzzuweisung zeigen sich keine besonderen Veränderungen gegenüber Simulation 1. Einzig im Fall der Reflexivpronomen zeigt sich ein etwas langsamerer Lernverlauf, der vermutlich auf die höhere Komplexität zurückzuführen ist. Für den Vergleich der pronominalen Formen nach Satztypen über die Epochen zeigt sich allerdings eine deutliche Änderung: Die komplexen Satztypen lassen sich im Falle des Personalpronomens nicht mehr so deutlich unterscheiden. Für die Reflexivpronomen nimmt

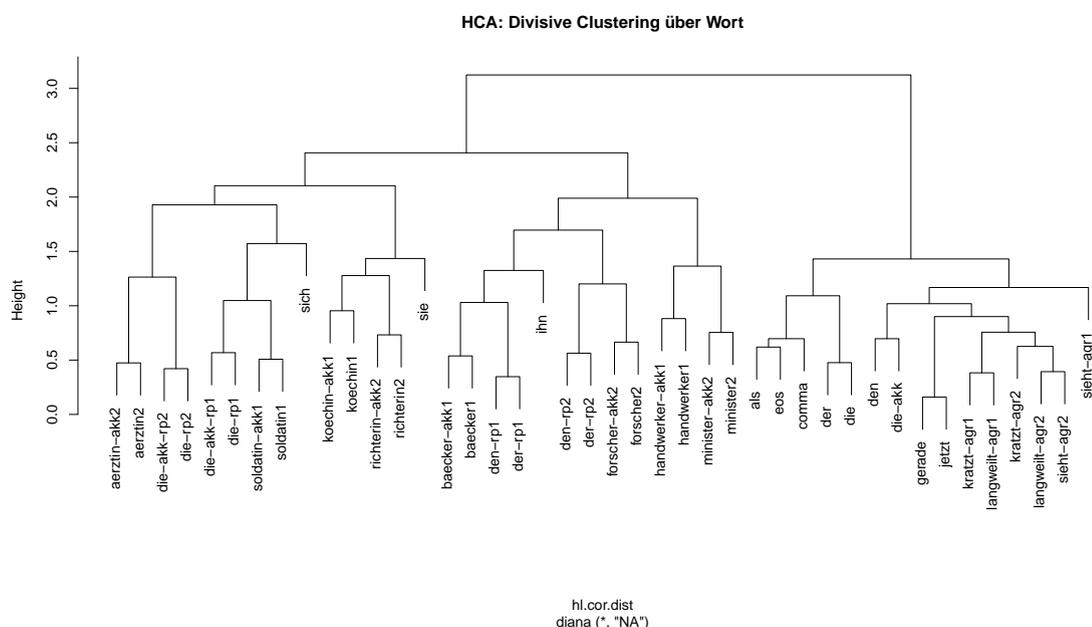


Abbildung 6.31.: Simulation 2. Hierarchische Cluster-Analyse (divisiv) der Hidden-Layer-Aktivierungen der Referenzzuweisungs-Komponente: Räumliche Gliederung der Wortformen.

die Unterscheidbarkeit zwischen komplexen Neben- und Hauptsätzen hingegen zu. In den Details der Aktivierung der Referentenklassen sind wieder kontextunabhängige Effekte für die Fehler bei der Referenzzuweisung verantwortlich, während LSK-Effekte ausgeschlossen werden können. Im Unterschied zu Simulation 1 weitet sich dieses Erklärungsmuster aber auf die Personalpronomen aus – für einen kontextuell bedingten Effekt durch die Distraktoren finden sich nur noch bedingt Hinweise. Folglich durchbricht das Erscheinen eines Kommas auch bei der Referenzzuweisung die Fortsetzung von LSK bzw. unterbindet deren Entstehen. Die Rolle der Interpunktion zur Unterbindung von Effekten des lokalen Kontexts werden wir in einer der folgenden Untersuchungen näher beleuchten. Allerdings ist an dieser Stelle darauf hinzuweisen, dass das Komma in unserer Simulation einen zu Wörtern gleichwertigen Status besitzt und damit die Erscheinungsform und Unterscheidbarkeit von Sequenzen maßgeblich verändert. Dies ist keine unproblematische und sicherlich auch keine realistische Annahme für die Rolle der Satzzeichen. Insofern ist der positive Effekt der Interpunktion nicht allzu verwunderlich und mit Vorsicht zu genießen.

In den Komponenten entwickeln sich wieder geeignete internen Repräsentationen für die jeweils zu erfüllenden Aufgaben. Außerdem finden sich keine Hinweise auf die Abbildung irrelevanter linearer Informationen. Insbesondere für die Repräsentation der Wortformen schlagen sich in der Wortvorhersage-Komponente die neuen grammatischen Markierungen nieder. Dabei zeigt sich, dass die Genus-Markierung in Simulation 1 in dieser Komponente nur in ihrer Funktion für die Prädikat-Kongruenz abgebildet wird, denn in Simulation 2 spielt sie keine Rolle für die Clusterung mehr. Dies ist aber kein Hindernis dafür, dass die Genus-Markierung in der Referenzzuweisungs-Komponente wieder den wichtigsten Gliederungsaspekt übernimmt. Das legt nahe, dass die internen Repräsentationen der

Wortvorhersage-Komponente nach wie vor genügend Informationen abbilden und weitergeben. Somit ermöglichen sie es der Referenzzuweisungs-Komponente, die nötigen Klassifizierungen des Inputs zu erfüllen. Letztere erzwingt also die Abbildung der nötigen Informationen in ersterer – ein Aspekt, auf den wir ebenfalls in einer der folgenden Untersuchungen nochmals eingehen werden.

### 6.5. Simulationsstudie 3: Studien mit reduzierten Stimuli

In drei vertiefenden Fragestellungen wollen wir weitere Eigenschaften unserer Architektur untersuchen: (a) Welche Rolle spielt die satzinterne Interpunktion für die unterschiedlichen Ergebnisse aus Simulation 1 und 2 und welche Effekte ergeben sich bei vorenthaltener Interpunktion? (b) Können Referenten auch dann korrekt als Antezedens erkannt werden, wenn sie während des Trainings nie als Antezedens einer entsprechenden Anapher aufgetreten sind? (c) Inwiefern sind die hier getesteten integrativen Netzwerke von einfachen SRNs zu unterscheiden, die ausschließlich Wortvorhersage leisten? Die Simulationen verbindet die gemeinsame Idee, die Leistungsfähigkeit und die Eigenschaften des Modells zu untersuchen, indem über eine Reduzierung der gegebenen Information verglichen wird, welche Aspekte der Leistungsfähigkeit und der Eigenschaften auf welchen Aspekten des Modells oder der Informationen beruhen. Wir erläutern und begründen in den nächsten Abschnitten kurz die gewählten Manipulationen.

**Korpus 3a: Vorenthaltene satzinterne Interpunktion** Die Variation der Grammatik zur Erzeugung von Korpus 3a besteht in der Auslassung aller satzinternen Satzstruktur-Markierungen: Es gibt zwar weiterhin die EOS-Markierung am Ende jedes vollständigen Satzes, die COMMA-Markierung zwischen Teilsätzen entfällt jedoch. Dies erfordert, dass abgeschlossene Abhängigkeiten innerhalb von Teilsätzen und Übergänge zwischen Teilsätzen ohne die zusätzlichen strukturellen Informationen erkannt werden. Wir konzentrieren uns in dieser Simulation auf Effekte dieser fehlenden Satzstruktur-Markierung: Zeigen sich im Vergleich zu Simulation 2 Unterschiede für den LSK-Effekt, wenn die LSK durch das Fehlen der Kommata nicht mehr durchbrochen werden? Zeigen sich eventuell deutlichere Effekte eines Komplexitätsunterschieds zwischen Subjekt- und Objekt-Relativsätzen? Hinsichtlich des problematischen Wort-Status des Kommas in Simulation 2 bietet die Variation in Simulation 3a einen vielleicht etwas realistischeren Blick auf mögliche Interferenz-Effekte. Unterschiede zwischen den beiden Simulationen geben aber möglicherweise auch einen Hinweis darauf, welche Funktionen Satzzeichen und Prosodie erfüllen können.

**Korpus 3b: Vorenthaltener Referent** Das Problem der Generalisierung über Sachverhalte, die im Trainings-Set nicht enthalten sind (z.B. Wörter, grammatische Funktionen, grammatische Zugehörigkeiten), wurde bereits in Kapitel 3.1.4 in der Diskussion zur Systematizität in SRNs erläutert, in der auch die Problematisierung durch G. F. Marcus (1998) dargestellt wurde. In Kapitel 6.1.1.1 haben wir angesprochen, wie dieser Aspekt in der Studie von R. Frank et al. (2005) auf die Frage der Anapherauflösung übertragen wurde. Im Kern geht es dabei um das Problem der Abbildung der Identitätsfunktion als Aspekt der Variablenbindung für unbekannte Werte: Wie systematisch ist der Erwerb und die

Leistung der Identitätsfunktion in SRNs bei Training mit unvollständiger Referenz. Bislang konnte nicht gezeigt werden, dass auf SRNs basierende Architekturen diese Funktion leisten können. Sollten SRNs die hierzu erforderliche Funktionalität nicht besitzen, ist dies durchaus eine bedeutsame Lücke im Erklärungsspektrum dieser Modellklasse. Woran liegt es, dass bislang keine Demonstration der erfolgreichen Bewältigung dieser Aufgabe in den bisherigen Studien mit SRNs vorgelegt werden konnte? Und wie können wir SRNs dazu bringen, sie doch zu bewältigen? Genügt es, Spezifika der Stimuli oder Spezifika der Architektur zu variieren? Wir besprechen in Kapitel 6.5.2 eine Reihe von Lösungsansätzen, die wir mit unserer Architektur getestet haben, und untersuchen die Merkmale der dabei erbrachten Performanz.

**Korpus 3c: Performanz ohne Referenzzuweisung** Die guten Performanzwerte bei der Wortvorhersage (v.a. bei der Auflösung von Fernabhängigkeiten bei der Subjekt-Prädikat-Kongruenz) legen die Möglichkeit eines positiven Aspekts der zusätzlichen Aufgabe der Referenzzuweisung auf die Performanz bei der Wortvorhersage nahe: Könnte es sein, dass die Referenzzuweisung im Sinne eines Bootstrapping-Mechanismus (vgl. z.B. Desai, 2007) durch die zusätzliche Information eine schnellere und bessere Ordnung der internen Struktur ermöglicht, die sich positiv auf die Leistung für die Wortvorhersage-Aufgabe auswirkt? Im zweiten Unterkapitel vergleichen wir die Performanz und die internen Repräsentationen der Netzwerke der vorausgegangenen Simulationen mit Netzwerken, die ohne Referenzzuweisungs-Komponente trainiert wurden. Dabei gilt es verschiedene Fragen zu beantworten: Wie gut und wie schnell lernen Netzwerke ohne zusätzliche Referenzzuweisungsaufgabe (also einfache SRNs) die Wortvorhersage für Korpus 2? Welche der beiden Netzwerk-Varianten schneidet besser ab (einfaches SRN vs. AnaSRN) bzw. gibt es überhaupt Unterschiede bei der Wortvorhersage-Performanz? Zeigen sich spezifische Unterschiede in der Performanz? Ergeben sich spezifische Unterschiede zwischen den beiden Netzwerk-Varianten für die interne Strukturierung der Hidden-Layer? Und falls es Unterschiede gibt: Wie lassen sich diese erklären? Wir besprechen eine Simulationsreihe mit einfachen SRNs und vergleichen die Merkmale mit den Ergebnissen aus der Simulationsstudie 2.

### 6.5.1. Korpus 3a: Effekte fehlender Satzstruktur-Markierung

Wie wirkt sich das Fehlen von satzinternen Interpunktionszeichen auf die Wortvorhersage und die Referenzzuweisung aus? Während das Netzwerk für Korpus 2 sehr gut gelernt hat, Teilsatzgrenzen zu unterscheiden und daraus Vorteile für die Wortvorhersage und die Referenzzuweisung zu ziehen, sollte das Netzwerk unseren Erwartungen zufolge größere Schwierigkeiten mit der Aufgabe haben, abgeschlossene Teilsätze zu erkennen, wenn diese Information wie in Korpus 3a nicht gegeben ist. Dies sollte sich auf die Wortvorhersage-Performanz auswirken, und zum Beispiel zu vermehrter Aktivierung falscher Wortklassen führen, die auf einer Fortsetzung lokaler Subsequenzen beruhen. In Simulation 2 konnten wir für die Referenzzuweisung eine allgemeine Fehlaktivierung für Referenten feststellen, die ähnliche grammatische Merkmale aufweisen wie das korrekte Antezedens. Allerdings schien dieser Effekt weitgehend unabhängig von Distraktoren des syntaktischen Kontexts zu erfolgen. Für die Verarbeitung von Korpus 3a erwarten

wir eine verstärkte Sensitivität für den syntaktischen Kontext und deutlicher erkennbare LSK-Effekte bei der Auflösung anaphorischer Referenzen. In den folgenden Abschnitten präsentieren wir die Ergebnisse der Simulationsstudie mit Korpus 3a. Die Grammatik ist identisch zur Version aus Simulation 2 mit Ausnahme des Fehlens der Kommata, die dementsprechend auch im Test-Korpus fehlen (zudem sind folglich die Input- und Output-Layer der Wortvorhersage-Komponente um eine Unit kleiner). Ansonsten ist auch das Test-Korpus identisch zu dem in Simulation 2 verwendeten (vg. Bsp. 24, Kap. 6.4.1). Korpus 3a umfasst insgesamt 325.892 Wörter, der längste Satz beinhaltet 115 Wörter, der kürzeste 5 Wörter, die durchschnittliche Satzlänge beträgt 16,3 Wörter. Alle technischen Spezifikationen und Details der Simulation entsprechen den Vorgaben aus Kapitel 6.3.2. Bei den Ergebnissen konzentrieren wir uns auf die Unterschiede zu den Ergebnissen für Korpus 2.

#### 6.5.1.1. Performanz bei der Wortvorhersage

Für die allgemeine Performanz zeigt sich das gewohnte Bild mit bereits zufriedenstellenden globalen Fehlerwerten nach Epoche 10 und nur noch marginalen Verbesserungen nach Epoche 30.<sup>49</sup> Wie gestaltet sich der globale Fehler über die verschiedenen Satzbedingungen? Betrachten wir zunächst den Satz in der LSK-Bedingung (Abb.6.32): Insgesamt lassen sich kaum Unterschiede zum Fehlerverlauf in Simulation 2 feststellen. Die einzigen Unterschiede finden sich auf den Positionen, auf denen sich in der ersten Simulation die Kommata befanden: So bleibt der Fehler auf dem Relativpronomen (entsprechend Position *comma\_1*) etwas höher als auf dem Relativpronomen in Simulation 1. Der Fehler auf der Anapher (entsprechend Position *comma\_2*) reduziert sich hingegen in den ersten Epochen etwas schneller. Und der Fehler auf dem finiten Verb des folgenden Hauptsatzes (entsprechend Position *comma\_3*) entwickelt sich deutlich langsamer und mit schlechterem Endresultat als in Simulation 2. Zum Vergleich ziehen wir die Kontrollbedingungen 2 (Abb. 6.33) und 4 (Abb. 6.34) heran (beide mit Einbettung eines Objekt-Relativsatzes)<sup>50</sup>: Zunächst können wir in Kontrollbedingung 2 einen massiven Fehler auf der Position des Relativsatz-Verbs feststellen, der in Simulation 2 nicht in diesem Umfang gegeben war. Auf der anschließenden Position der Anapher findet sich im Vergleich zur LSK-Bedingung eine etwas langsamere Fehlerverbesserung. Zwischen dem globalen Fehler der LSK-Bedingung und der Kontrollbedingung 2 zeigen sich jedoch keine Unterschiede, die auf einen LSK-Effekt hinweisen würden. Für die Kontrollbedingung 4 ergibt sich im Unterschied zum Pendant in Simulation 2 sowie im Unterschied der Bedingungen mit vorangestelltem Nebensatz in Simulation 2 eine etwas schnellere und konstantere Entwicklung des Fehlers auf dem Verb des Matrixsatzes.

Untersuchen wir zunächst, woher der hohe Fehler auf dem Relativsatz-Verb in Kontrollbedingung 2 stammt: Neben der Aktivierung von Verben der falschen Kongruenz-Klasse beruht der Fehler vor allem in der Aktivierung der Artikel sowie der Personal- und Re-

<sup>49</sup>Epoche 10:  $\bar{x}GPE_n = 0.157$  (Standardabweichung  $\sigma = 0.103$ ), auf die Verben bezogener Fehler  $\bar{x}GPE_n = 0.197$  (Standardabweichung  $\sigma = 0.099$ ). Epoche 30:  $\bar{x}GPE_n = 0.119$  (Standardabweichung  $\sigma = 0.069$ ), auf die Verben bezogener Fehler  $\bar{x}GPE_n = 0.129$  (Standardabweichung  $\sigma = 0.053$ ).

<sup>50</sup>Die Fehlerverläufe für die Sätze mit eingebettetem Subjekt-Relativsatz sind weitgehend identisch zu ihren jeweiligen Pendants.

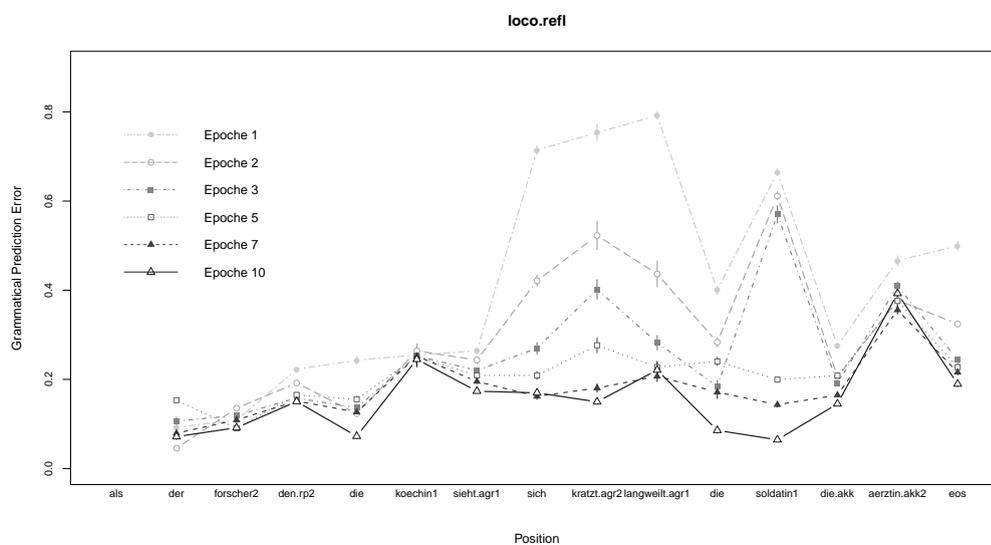


Abbildung 6.32.: Simulation 3a. Verlauf und Entwicklung des globalen Fehlers der Wortvorhersage-Komponente, gruppiert nach Epoche: GPE über Wortpositionen für Nebensätze mit zentral eingebettetem Objekt-Relativsatz (LSK-Bedingung).

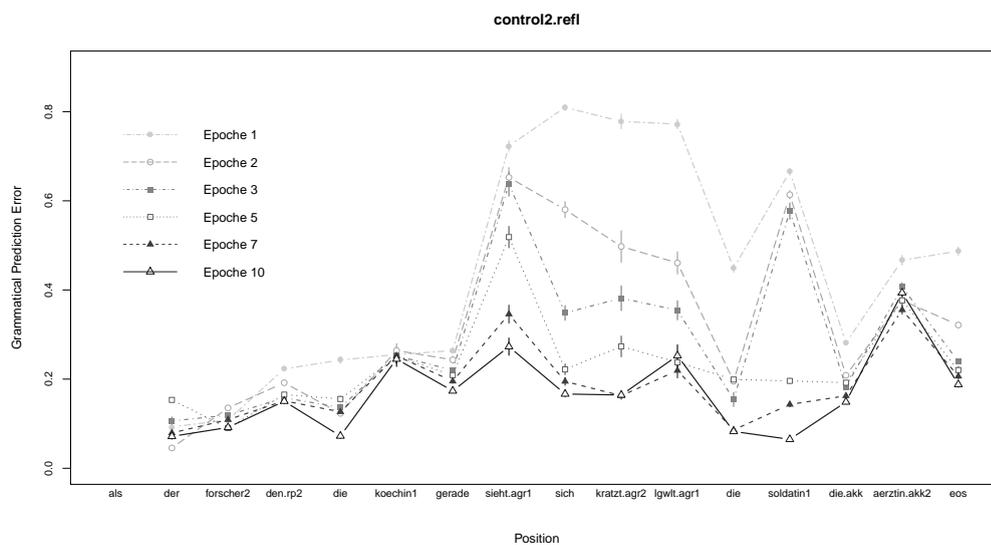


Abbildung 6.33.: Simulation 3a. Verlauf und Entwicklung des globalen Fehlers der Wortvorhersage-Komponente, gruppiert nach Epoche: GPE über Wortpositionen für Nebensätze mit zentral eingebettetem Objekt-Relativsatz und eingefügtem Adverb (Kontrollbedingung 2).

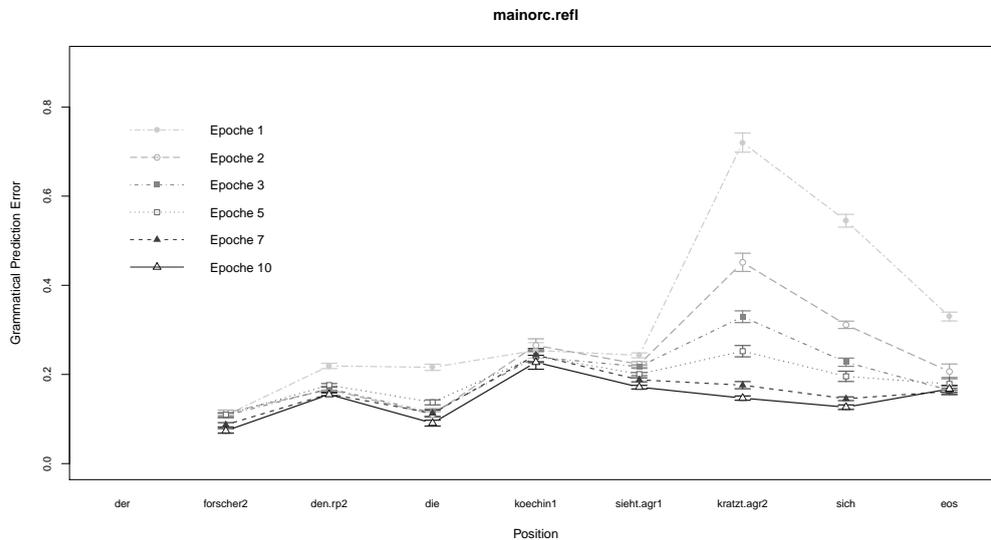


Abbildung 6.34.: Simulation 3a. Verlauf und Entwicklung des globalen Fehlers der Wortvorhersage-Komponente, gruppiert nach Epoche: GPEn über Wortpositionen für Hauptsätze mit zentral eingebettetem Objekt-Relativsatz (Kontrollbedingung 4).

flexivpronomen. Dies läßt sich nur dadurch erklären, dass das vorhergehende Adverb als finites Verb aufgefasst wird. Dies ist hinsichtlich seiner Kollokation nicht unerwartet: Das Adverb hat in unserer künstlichen Grammatik ein ähnliches Umfeld wie die Verben, im Unterschied zur Wortfolge in Hauptsätzen erscheint es dabei in Nebensätzen immer vor dem Verb. Es nimmt in Kontrollbedingung 2 also die Position ein, die das Verb im Hauptsatz einnehmen würde. Folglich muss in direkter Folge entweder der Artikel des Akkusativ-Objekts oder ein Pronomen folgen. Aus dieser Perspektive ist die Verwendung der Kontrollbedingung 2 kein geeigneter Kontrast zur LSK-Bedingung, denn das Adverb führt in den frühen Epochen zu einem Frequenz-Effekt durch die unvollständig erworbene Differenzierung zwischen Adverbien und Verben.

Zum Vergleich der LSK-Bedingung mit Kontrollbedingung 4 (Hauptsatz mit eingebettetem Objekt-Relativsatz) ziehen wir die Positionen des Verbs und des Pronomens im Matrixsatz heran. Dabei stimmen zuerst die Positionen des Pronomens in der LSK-Bedingung (Abb. 6.35) und des Matrixsatz-Verbs in der Kontrollbedingung 4 (Abb. 6.36) überein. Der größte Anteil des Fehlers in der LSK-Bedingung ist durch die Aktivierung der Verben bedingt. Die direkte Abfolge zweier Verben ist nur im Übergang von einem Relativsatz möglich, der in einen Hauptsatz eingebettet ist (oder alternativ im Übergang zwischen einem vorangestellten Nebensatz zu einem Hauptsatz mit entsprechender Topikalisierung des Verbs). Damit zeigt sich hier weniger ein LSK-Effekt als vielmehr ein Frequenz-Regularitäts-Effekt, den wir mittlerweile schon des öfteren beobachten konnten. Für Kontrollbedingung 4 zeigt sich neben der Fehlaktivierung von Verben der falschen Kongruenz-Klasse in den frühen Epochen eine Aktivierung von Artikeln und Adverbien. Gerade dieser Fehler entspricht unseren Erwartungen eines LSK-Effekts: Nur wenn das vorausgegangene Verb als finites Verb eines Hauptsatzes erfasst wurde, kann in der Folge ein Adverb oder ein Artikel aktiviert werden.

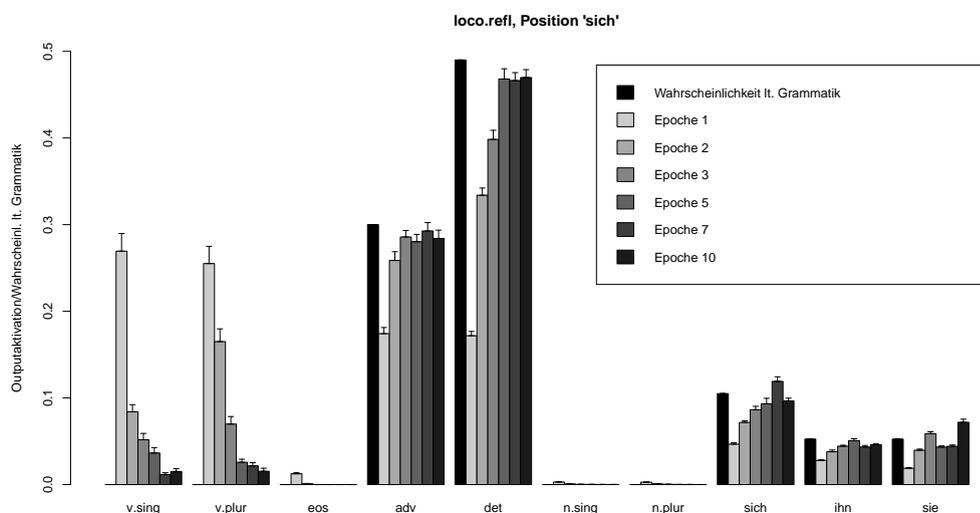


Abbildung 6.35.: Simulation 3a. Aktivierung für die Anapher: Details der Aktivierung im Vergleich zur absoluten Wahrscheinlichkeit der Aktivierungsverteilung in vorangestellten Nebensätzen nach eingebettetem Objekt-Relativsatz (*Als der Forscher den der Bäcker sieht sich kratzt ...*).

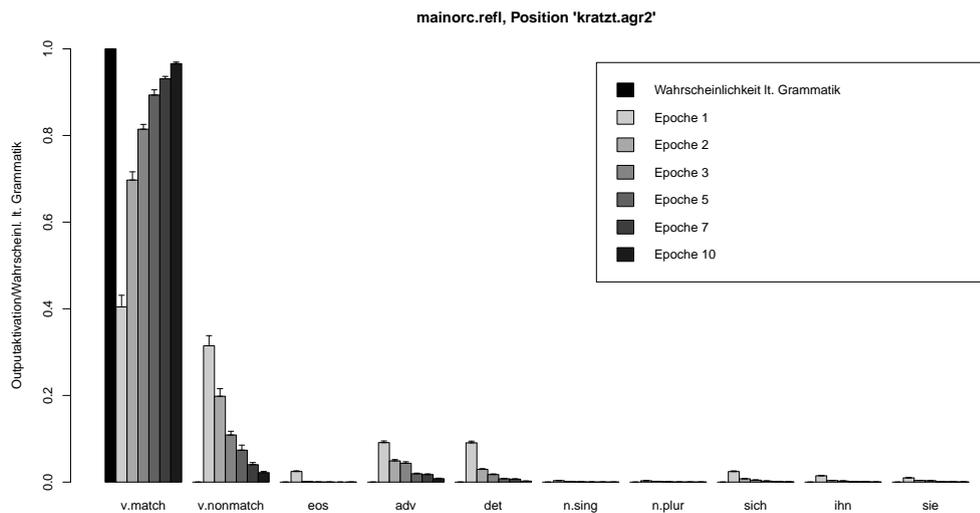


Abbildung 6.36.: Simulation 3a. Aktivierung für das Verb des Matrixsatzes: Details der Aktivierungsverteilung in Hauptsätzen nach eingebettetem Objekt-Relativsatz (*Der Forscher den der Bäcker sieht kratzt sich.*).

Auf der anschließenden Position stehen sich in der LSK-Bedingung das Matrixsatz-Verb (Abb. 6.37) und in der Kontrollbedingung die Anapher (Abb. 6.38) gegenüber: In der LSK-Bedingung findet sich neben der üblichen Fehlaktivierung der Verben der falschen Kongruenz-Klasse eine kurzzeitige Aktivierung des EOS-Markers sowie der Adverbien. Die Aktivierung des EOS-Markers entspricht wiederum unseren Erwartungen eines LSK-Effekts: Das Subjekt und das Verb des Relativsatzes ergeben mit der Anapher lokal einen vollständigen Hauptsatz (...*die Köchin sieht sich*...). Folglich müsste anstelle des Matrixsatz-Verbs ein EOS vorhergesagt werden. Die Tatsache, dass die gleiche Aktivierung des EOS-Markers auch in Kontrollbedingung 1 (nach eingebettetem Subjekt-Relativsatz) erscheint, weist allerdings darauf hin, dass das Netzwerk nicht sensitiv gegenüber der Kasus-Markierung zu sein scheint. Die Aktivierung der Adverbien schließlich ist schwer zu erklären. Eine Möglichkeit besteht darin, dass sich die vorhergehende Fehlaktivierung der Verben im Sinne einer Hauptsatzfortführung nach eingebettetem Relativsatz auf die Folgeaktivierung auswirkt, denn nur nach einem Verb im regulären Hauptsatz kann ein Adverb folgen. Eine weitere Möglichkeit wäre, dass in den frühen Epochen noch nicht hinreichend zwischen Pronomen und Verben unterschieden wird, und deshalb nach dem vermeintlichen Verb die Adverbien aktiviert werden. Aber diese Möglichkeiten sind reine Spekulation. In Kontrollbedingung 4 finden sich vor allem eine frühzeitig reduzierte Fehlaktivierung der Verben. Hierfür lässt sich allerdings nicht einmal bei einer Übertragung der zuvor gefundenen Fehlaktivierungen eine schlüssige Erklärung liefern.

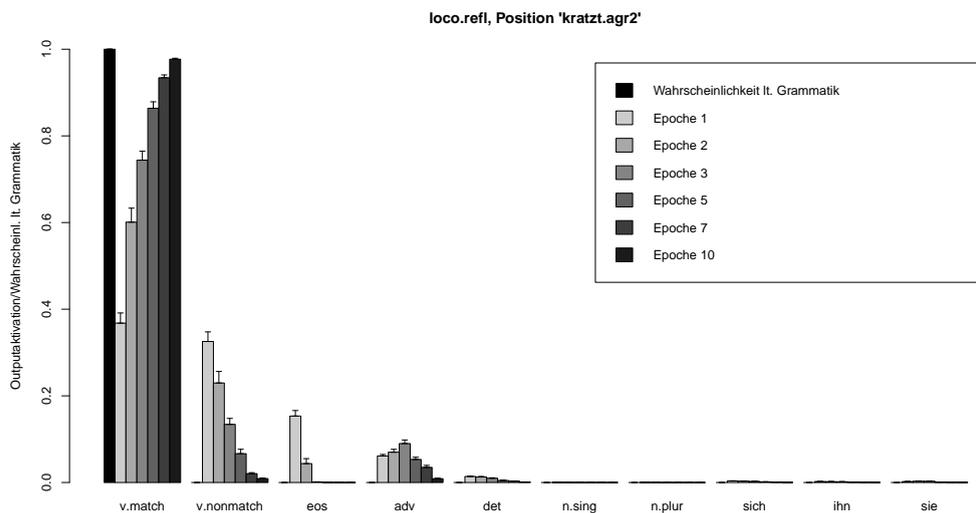


Abbildung 6.37.: Simulation 3a. Aktivierung für das Verb des Matrixsatzes: Details der Aktivierung im Vergleich zur absoluten Wahrscheinlichkeit der Aktivierungsverteilung in vorangestellten Nebensätzen nach eingebettetem Objekt-Relativsatz (*Als der Forscher den der Bäcker sieht sich kratzt ...*).

Der Fehler auf dem Relativpronomen sowie auf dem Verb des folgenden Hauptsatzes scheint genau die Fehlleistungen zu vertreten, die wir zuvor in Simulation 2 auf den Positionen des ersten und dritten Kommas gefunden haben: Auf dem Relativpronomen findet sich in vorangestellten Nebensätzen eine leicht überhöhte Aktivierung der Artikel im Akkusativ, in Hauptsätzen eine leicht überhöhte Aktivierung des Verbs. Beide Aktivierungen spiegeln

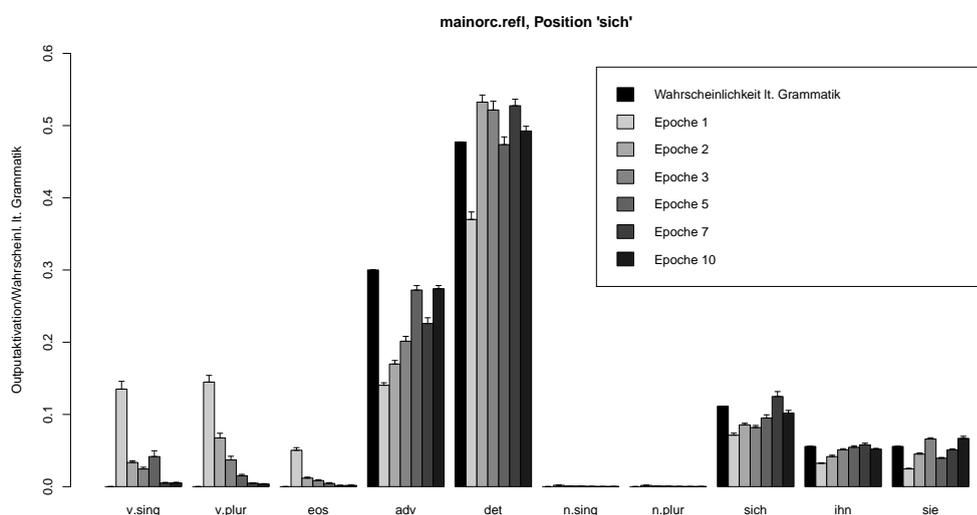


Abbildung 6.38.: Simulation 3a. Aktivierung für die Anapher: Details der Aktivierung im Vergleich zur absoluten Wahrscheinlichkeit der Aktivierungsverteilung in Hauptsätzen nach eingebettetem Objekt-Relativsatz (*Der Forscher den der Bäcker sieht kratzt sich.*).

die Orientierung an der Fortsetzung eines regulären Haupt- oder Nebensatzes ohne Relativsatzeinbettung wider. Die langsame Fehlerentwicklung auf dem Hauptsatzverb beruht auf der falschen Aktivierung des EOS-Markers, die wir in Simulation 2 auf dem Komma als Hinweis auf die Erwartung eines abgeschlossenen Nebensatzes gefunden haben.

### 6.5.1.2. Performanz bei der Referenzzuweisung

Hinsichtlich des globalen Fehlers und der allgemeinen Performanz zeigt sich auch für die Referenzzuweisungs-Komponente das gewohnte Bild: Erst ab Epoche 80 stabilisiert sich die Entwicklung der relevanten Anteile des globalen Fehlers auf einem gleichbleibenden Niveau.<sup>51</sup>

**Globale Analyse** Betrachten wir zunächst den globalen Fehler der Referenzzuweisung auf der Position der Anapher, gruppiert nach den verschiedenen Satztypen in ihrer Entwicklung über die Epochen: Für die Reflexivpronomen (Abb. 6.39) zeigt sich im Vergleich zu den Ergebnissen aus Simulation 2 eine wesentlich schnellere Verbesserung der Performanz für die komplexen Sätze. Als vergleichsweise schlechteste Bedingung etabliert sich auf den letzten Epochen Kontrollbedingung 1 (Nebensatz mit eingebettetem Subjekt-Relativsatz). Die Entwicklung der Personalpronomen (Abb.6.40) verläuft hingegen weitgehend ähnlich wie in Simulation 2. Im Unterschied dazu grenzen sich allerdings die Bedingungen mit

<sup>51</sup>Der allgemeine Zuweisungsfehler erhält hier einen Mittelwert von  $\bar{x}RAEn = 0.013$  (Standardabweichung  $\sigma = 0.056$ ). Der Fehlerwert für alle pronominalen Formen beläuft sich auf  $\bar{x}RAEn = 0.08$  (Standardabweichung  $\sigma = 0.121$ ). Der ausschließlich auf Reflexiv- und Personalpronomen bezogene Fehlerwert beläuft sich auf  $\bar{x}RAEn = 0.152$  (Standardabweichung  $\sigma = 0.13$ ).

Nebensatz beginnend ab Epoche 50 etwas deutlicher von den anderen Bedingungen ab (v.a. die LSK-Bedingung und die Kontrollbedingung 1).

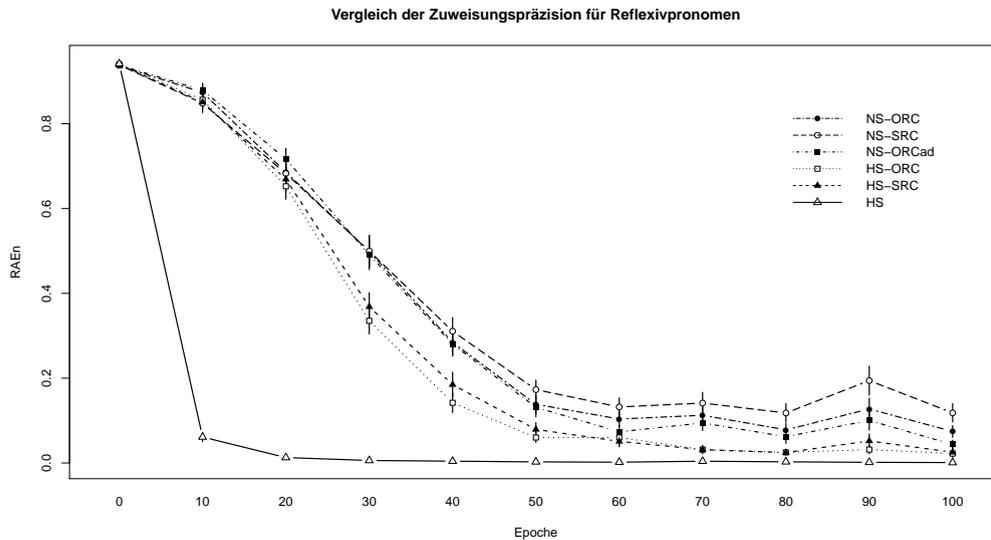


Abbildung 6.39.: Simulation 3a. Vergleich der RAEn für Reflexivpronomen in verschiedenen Satztypen über den Verlauf der Epochen (Entwicklung).

In Abbildung 6.41 vergleichen wir wieder die Anapherntypen über die verschiedenen Satztypen hinweg in einem Querschnitt über mehrere Epochen mit fortgeschrittener konstanter Performanz (Epoche 50 bis Epoche 100). Während die Performanz und die Relationen über die verschiedenen Satztypen bei den Personalpronomen weitgehend mit den Ergebnissen aus Simulation 2 übereinstimmen, zeichnet sich für die Reflexivpronomen eine leichte Veränderung ab: Hier kann eine insgesamt bessere Performanz für alle Satztypen außer den einfachen Hauptsätzen festgestellt werden. Der stärkste Unterschied in den Relationen zeigt sich für die Kontrollbedingung 2 (Anapher nach Objekt-Relativsatz mit eingefügtem Adverb): Diese Bedingung wird sowohl bei den Reflexiv- als auch bei den Personalpronomen deutlich besser verarbeitet als die LSK-Bedingung und die Kontrollbedingung 1. Wie wirken sich die Unterschiede der Performanz im Detail aus und auf welche Ursache sind sie zurückzuführen? Wir vergleichen wieder die gemittelten Aktivierungen über die Referentenklassen aus Simulation 2.

**Detailanalyse: Reflexivpronomen** Abbildung 6.42 zeigt die Aktivierungen für die Reflexivpronomen, wobei wir uns auf die ersten 60 Epochen beschränken. Beginnen wir mit dem Antezedens: Wie in Simulation 2 werden die Referenten in einfachen Hauptsätzen ab Epoche 10 weitgehend fehlerfrei aktiviert, während die korrekten Referenten in den komplexen Satztypen erst ab Epoche 30 stärker aktiviert werden als die Distraktoren und andere unzulässige Antezedenzen. Bis Epoche 30 unterscheiden sich einfache Hauptsätze signifikant von komplexen Sätzen. In den folgenden Epochen verliert sich dieser Unterschied aber sukzessive: Ab Epoche 40 unterscheiden sich einfache Hauptsätze nicht mehr signifikant von Hauptsätzen mit Subjekt-Relativsatz, ab Epoche 50 nicht mehr von beiden komplexen Hauptsatztypen und ab Epoche 60 nicht mehr von Kontrollbedingung

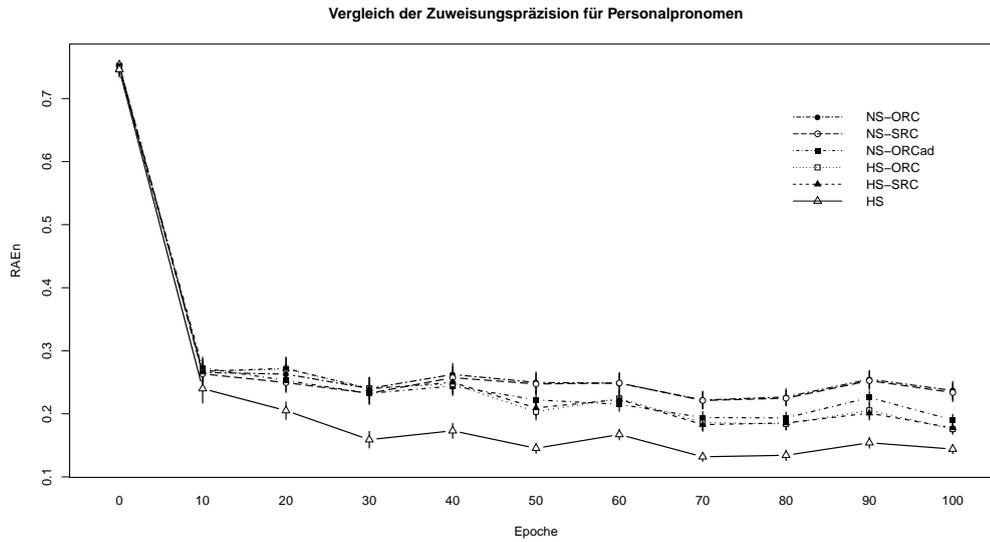


Abbildung 6.40.: Simulation 3a. Vergleich der RAEn für Personalpronomen in verschiedenen Satztypen über den Verlauf der Epochen (Entwicklung).

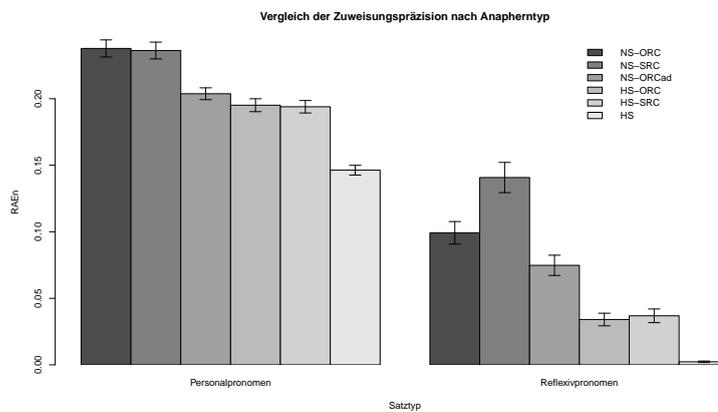


Abbildung 6.41.: Simulation 3a. Vergleich des RAEn für verschiedene anaphorische Formen (Reflexivpronomen und Personalpronomen), gemittelt über die Epochen 60-100, gruppiert nach Satztypen (sortiert von komplex zu einfach).

2 (Nebensatz mit Adverb im Relativsatz). Ab Epoche 30 zeigt sich auch ein Unterschied für die komplexen Satztypen: Die Aktivierung der Referenten in den Nebensätzen ist für alle drei Satztypen gegenüber den Satztypen mit Hauptsatz signifikant geringer. Dieser Unterschied wird über die weiteren Epochen ebenfalls sukzessive nivelliert, wobei sich der Unterschied zwischen Kontrollbedingung 1 (Nebensatz mit Subjekt-Relativsatz) gegenüber den Hauptsatz-Strukturen am längsten hält.

Betrachten wir die Distraktoren und die alternativen Referenten: Wie zuvor zeichnet sich auch in Simulation 3a wieder ab, dass die alternativen Aktivierungen vor allem durch die Klasse der Referenten bestimmt werden, die sowohl im Genus als auch in der Prädikat-Kongruenz mit dem korrekten Antezedens übereinstimmen. In den anderen Referentenklassen werden vermutlich nur die Aktivierungsverhältnisse reflektiert, die in dieser Klasse mit vollständig übereinstimmender Kongruenz zugegen sind.<sup>52</sup> Das zeigt sich auch darin, dass ab Epoche 20 alle anderen Klassen (außer der des korrekten Antezedens) signifikant schwächer aktiviert werden als diese. Deshalb beschränken wir uns im Folgenden auf die Klasse 4\_VerbGenK: Der Unterschied der geringeren (bzw. praktisch nicht vorhandenen) Aktivierung falscher Referenten bei einfachen Hauptsätzen gegenüber den komplexen Sätzen bleibt bis Epoche 40 signifikant. Ab Epoche 50 verliert er sich zuerst gegenüber den komplexen Hauptsätzen und Kontrollbedingung 2, ab Epoche 60 gegenüber allen komplexen Satzstrukturen. Zwischen den komplexen Satzstrukturen zeigen sich nur selten Unterschiede – in Epoche 30 für die LSK-Bedingung gegenüber Hauptsätzen mit Subjekt-Relativsatz, in Epoche 50 für Kontrollbedingung 1 gegenüber beiden komplexen Hauptsatz-Strukturen (dieser Effekt bleibt bis Epoche 100 erhalten). Die Befunde sprechen wieder für einen allgemeinen kontextunabhängigen Interferenz-Effekt beim Abruf der relevanten Informationen in Interaktion mit der unterschiedlichen Komplexität der verschiedenen Satztypen. Allerdings scheint das Adverb in Kontrollbedingung 2 einen (wenn auch geringfügigen) positiven Einfluss auf die Performanz zu nehmen, was auf einen leichten Einfluss des syntaktischen Kontexts hindeutet.

**Detailanalyse: Personalpronomen** Abbildung 6.43 zeigt die Aktivierungen für die Personalpronomen für die Epochen 10 bis 80. Die Aktivierung der korrekten Antezedenzen ist ab Epoche 20 signifikant besser in einfachen Hauptsätzen als in den anderen Satzstrukturen: In diesen Sätzen verbessert sich die korrekte Zuweisung schneller als in den komplexen Sätzen. Zwischen den komplexen Sätzen gibt es hingegen über alle Epochen hinweg keine signifikanten Unterschiede. Interessanter ist der Verlauf in der Referenten-Klasse der Distraktoren: Ab Epoche 40 wird der entsprechende Referent in der LSK-Bedingung signifikant weniger stark aktiviert als in den komplexen Hauptsätzen. Ab Epoche 50 trifft dies sowohl auf die LSK-Bedingung als auch auf die Kontrollbedingung 1 zu. Und ab Epoche 60 unterscheiden sich diese beiden Bedingungen zusätzlich von Kontrollbedingung 2 (Nebensatz mit Adverb im Relativsatz). Dies ist genau der Effekt, den wir erwartet haben und der sich ausschließlich im Sinne einer LSK erklären lässt: Da der Distraktor im lokalen

<sup>52</sup>Um diesen Gedankengang nochmals zu verdeutlichen: Die Aktivierung der Klassen der allgemeinen Referenten – d.h. die Referenten, die unabhängig von ihrem Auftreten im jeweiligen Satz betrachtet werden – enthält natürlich auch die Aktivierungen der Distraktoren. Wenn die eingeschränktere Klasse mit vollständiger Kongruenzübereinstimmung stärker aktiviert ist als die weniger restriktiven Klassen, dann ist anzunehmen, dass die berücksichtigten Merkmale unmittelbar relevant sind und mehr Einflussfaktoren offenlegen können.

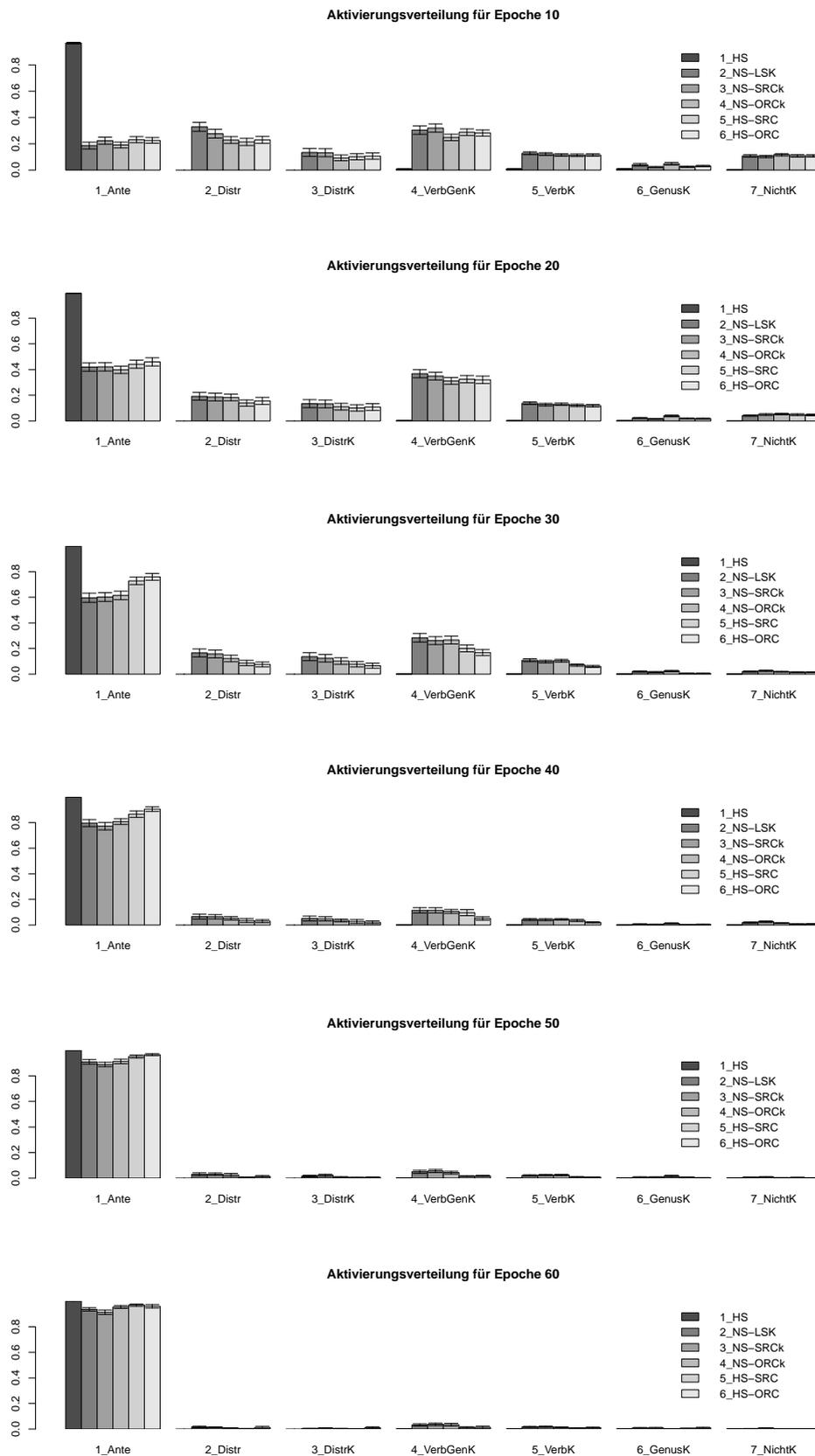


Abbildung 6.42.: Simulation 3a. Referenten-Aktivierung für Reflexivpronomen.

Kontext die Position des Matrixsatz-Subjekts einnimmt, wird er als mögliches Antezedens des Personalpronomens tendenziell ausgeschlossen. Dass sich der Effekt auch für den Nebensatz mit Subjekt-Relativsatz zeigt, weist allerdings darauf hin, dass das Netzwerk hierbei nicht auf die Kasus-Information achtet (das vermeintliche lokale Satz-Subjekt hat hier den Kasus Akkusativ). Die gleichen Effekte deuten sich umgekehrt bei der Aktivierung des Matrixsatz-Subjekts an. Allerdings wird der Unterschied mit einer stärkeren Aktivierung des falschen Referenten in der LSK-Bedingung und der Kontrollbedingung 1 erst in Epoche 100 signifikant.

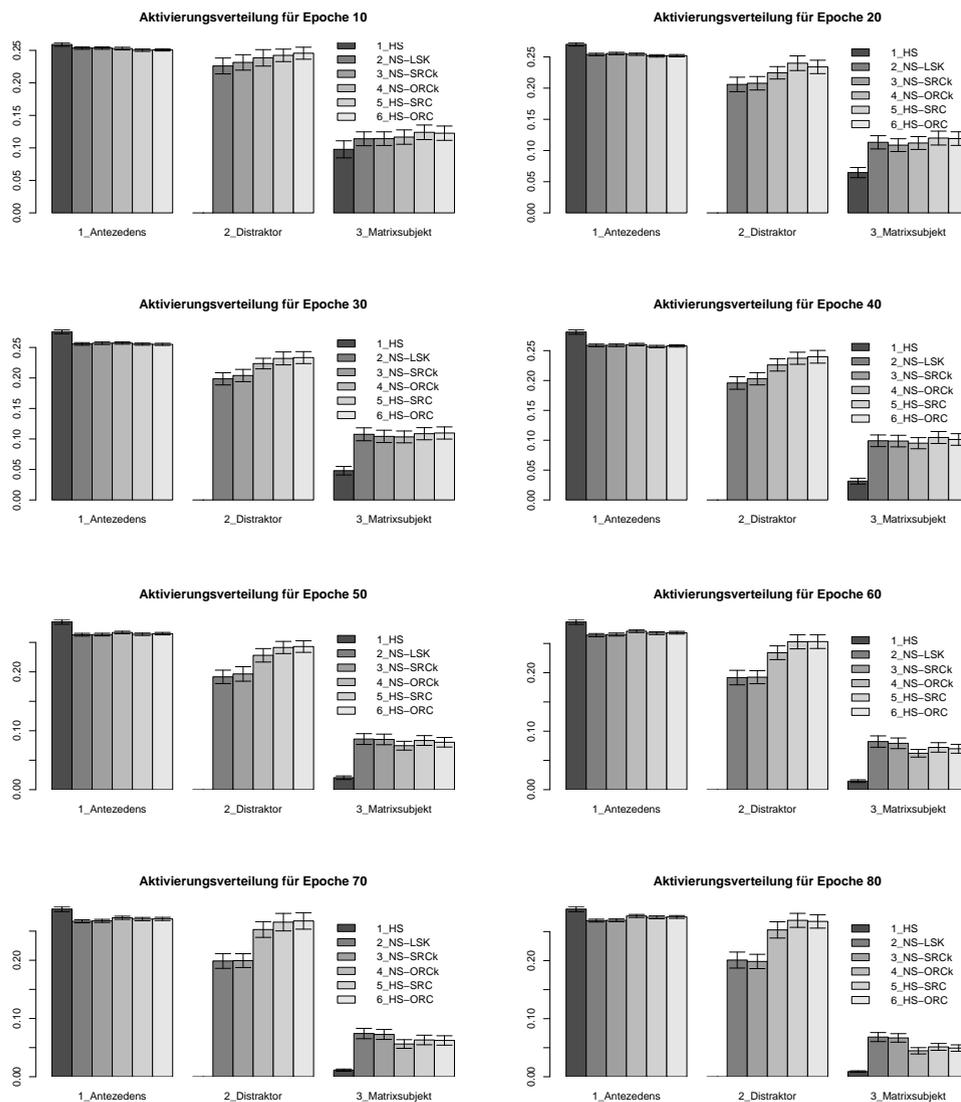


Abbildung 6.43.: Simulation 3a. Details der Referenten-Aktivierung für Personalpronomen, Epochen 10 bis 80.

### 6.5.1.3. Diskussion

Die fehlende Satzstruktur-Markierung wirkt sich auf die Wortvorhersage auf verschiedenen Positionen mit vermehrten Fehlaktivierungen aus, die einem LSK-Effekt zugeschrieben werden können. Vor allem die für die LSK-Bedingung und die Kontrollbedingung 4 gefundenen Fehler entsprechen unseren diesbezüglichen Erwartungen. Allerdings finden sich entsprechende Effekte auch in den Bedingungen mit Subjekt-Relativsatz. Wir nehmen den ironischen Sachverhalt zur Kenntnis, dass sich damit auch ausgerechnet in Satzbedingungen entsprechende Effekte finden, die wir als Kontrollbedingungen konzipiert hatten. Das Netzwerk scheint demzufolge auf den entsprechenden Positionen Kasus-Informationen zu ignorieren, die LSK-Effekte verhindern müssten, da das vermeintliche Subjekt der lokalen Sequenz in diesem Falle den Kasus Akkusativ aufweist. Eine mögliche Erklärung könnte die folgende sein: Auf den Positionen, auf denen ein Akkusativ-Objekt erscheint, sind die Informationen über das Satz-Subjekt weitgehend irrelevant, da dessen grammatische Merkmale (Prädikat-Kongruenz, Genus) keine Bedingungen an die Eigenschaften des Akkusativ-Objekts stellen. Deshalb könnte auf der entsprechenden Position, auf der der LSK-Effekt erscheint, die relevante Information nicht präsent genug sein, dass das vermeintliche lokale Subjekt nicht über den Kasus Nominativ verfügt. Diese Erklärung ist allerdings sehr spekulativ.

Auch für die Referenzzuweisung finden wir wieder verstärkt LSK-Effekte, die sich vor allem auf die Auflösung der Personalpronomen auswirken. Für diese finden sich die erwarteten Effekte: für die Klasse des Distraktors der Effekt der reduzierten Aktivierung, für das Matrixsatz-Subjekt der Effekt der verstärkten falschen Aktivierung. Dabei scheint für die Sätze in Kontrollbedingung 1 wiederum die Kasus-Information ignoriert zu werden, denn das lokale Nomen trägt mit dem Akkusativ den falschen Kasus, um als potenzielles Subjekt der lokalen Subsequenz und damit als nicht zugängliches Antezedens interpretiert zu werden. Allerdings unterscheiden sich die betroffenen Bedingungen von den Kontrollbedingungen 2, 3 und 4, in denen die LSK alleine schon durch die Wortabfolge durchbrochen wird. Der LSK-Effekt scheint demnach vor allem von den Regularitäten der Wortabfolge abzuhängen. Bei den Reflexivpronomen zeigt sich wiederum ein kontextunabhängiger Interferenz-Effekt, der mit der unterschiedlichen Komplexität der verschiedenen Satztypen interagiert. Die geringfügig niedrigeren Fehlerwerte in Kontrollbedingung 2, die ja eigentlich noch komplexer sein sollte, deuten allerdings auf einen leichten Einfluss des syntaktischen Kontexts hin, der durch das Adverb und die damit durchbrochene LSK erklärt werden könnte.

### 6.5.2. Korpus 3b: Vorenthaltener Referent

In unserer Studie zur Frage des Erwerbs der Identitätsfunktion außerhalb des Trainingspektrums verwenden wir eine Variation unserer bisherigen Trainingsstimuli, in der zwei ausgewählte Referenten nie als Antezedens des Reflexivpronomens erscheinen. Im Unterschied zur „*rose-is-a-rose*“-Studie (G. F. Marcus, 1998), in der ein gänzlich neues Wort gebunden werden sollte, erscheinen die beiden Referenten während des Trainings ganz normal als explizit benanntes Subjekt oder Objekt in beliebigen syntaktischen Kontexten sowie als Antezedens von Relativpronomen und Personalpronomen. Die Referenten sind

also aus dem Training bekannt. Für die Input-Wörter, die auf diese Referenten verweisen, sollten also alle relevanten grammatischen und referenziellen Eigenschaften erworben worden sein. Ausschließlich die Funktion als möglicher Referent des Reflexivpronomens ist ausgeschlossen. Dies entspricht der Situation, die G. F. Marcus (1998, Seite 269) als außerhalb des Training-Spektrums „with respect to some particular function“ bezeichnet (vgl. Kap. 3.1.4).

Wir modellieren den Gegenstand in unserer Architektur mit integriertem Training von Wortvorhersage und Referenzzuweisung. In dieser Architektur gibt es im Unterschied zu den Simulationsstudien von G. F. Marcus (1998) eine klare Trennung von Wortform und Wortbedeutung (Referenz). Während in den früheren Studien eher eine Bindung von Variablen an Variablen simuliert wurde, entspricht unsere Simulation eher einer Bindung von veränderlichen Werten (d.h. Inhalten) an Variablen. Zwar entspricht die Anapherlösung formell eher der ersten Version, denn es wird ja nur die koreferierende Wortform gesucht. Dennoch erfolgt dadurch auch ein Verweis auf einen Referenten, und es ist denkbar, dass wir im natürlichsprachlichen Spracherwerb gerade über diese Referenzialität und die Situiertheit des Spracherwerbs die Variablenbindungs- und Identitätsfunktion erwerben. Da die allgemeine funktionale Verwendung der ausgewählten Referenten im Training erworben werden kann und nur ihre Bindung an Reflexivpronomen funktional außerhalb des Trainingspektrums liegt, besteht die Möglichkeit, dass die inhaltliche Zuweisung der Referenten als Satz-Objekt der entsprechenden Satzposition gelingt. Gleich vorab sei gesagt, dass auch unser Ansatz nicht die gewünschten Ergebnisse erbrachte. Dennoch müssen auch Misserfolge diskutiert werden, um Aspekte des Problems besser zu verstehen und um zu sehen, welche Ansätze sich nicht zur Lösung eignen. Abgesehen davon ermöglicht auch die nicht erfolgreiche Erfüllung dieser Aufgabe Einblicke in die Funktionsweise der hier untersuchten Architektur. Wir besprechen zuerst den allgemeinen Lösungsansatz und erläutern in einem zweiten Abschnitt in verkürzter Form weitere Versuche, um potenzielle Hindernisse des Erwerbs zu überwinden.

Für die Simulationen wurde Grammatik 2 als Ausgangspunkt verwendet. Außer der Anpassung der Constraints zur Auslassung der Verwendung der ausgewählten Referenten mit Reflexivpronomen wurde für Korpus 3b aber nichts verändert. Über die Constraints wurde ein Referent mit Genus Maskulinum mit Kongruenz-Markierung *agr1* (*Bäcker*) sowie eine Referentin mit Genus Femininum mit Kongruenz-Markierung *agr2* (*Ärztin*) von den möglichen Antezedenzien der Reflexivpronomen ausgenommen. Auf diese Weise können wir alle Eigenschaften, Fehlerursachen und Einflussfaktoren bei der Referenzzuweisung überprüfen. Dementsprechend wurde das Test-Set angepasst: Im Unterschied zur Version in den vorausgehenden Simulationen enthält die hier verwendete Version zwar weiterhin alle Satztypen. Allerdings verwenden wir nur noch die Versionen mit Reflexivpronomen und hierfür eine vermehrte Anzahl von Sätzen für die beiden ausgewählten Referenten als möglichen Antezedenzien. Die technischen Spezifika der Simulation entsprechen wieder den Vorgaben aus Simulation 2.

### 6.5.2.1. Ergebnisse zur Systematizität der Referenzzuweisung

Für die Feststellung der allgemeinen Fähigkeit zur korrekten Bindung der Referenten an das Reflexivpronomen genügt die Überprüfung der Performanz in einfachen Hauptsätzen.

Wird nicht einmal hier eine entsprechende Aktivierung entdeckt, so müssen wir davon ausgehen, dass die Aufgabe allgemein nicht bewältigt und die entsprechende Abstraktion nicht gelernt wurde. Abbildung 6.44 zeigt den globalen Fehler für einfache Hauptsätze. Das Diagramm läßt keine Fragen offen: Die für die korrekte Zuweisung des Reflexivpronomens zum ausgelassenen Referenten erforderliche Abstraktion wurde eindeutig nicht gelernt.<sup>53</sup> Die Aktivierungsdetails zu diesem Satz für die Position des Reflexivpronomens bei maskulinem Antezedens zeigen, dass die beiden ausgenommenen Referenten nicht, ansonsten aber ausnahmslos alle Referenten unabhängig ihres Genus oder ihrer Prädikat-Kongruenz aktiviert werden. Verweist das Matrixsatz-Subjekt hingegen auf einen der anderen Referenten, der folglich das Antezedens des Reflexivpronomens ist, werden wie in Simulation 2 nur diejenigen Referenten längerfristig mitaktiviert, die eine Übereinstimmung der Prädikat-Kongruenz aufweisen. Einzige Ausnahme bilden auch hier die beiden ausgenommenen Referenten, die selbst bei Übereinstimmung der Kongruenz nicht mitaktiviert werden.

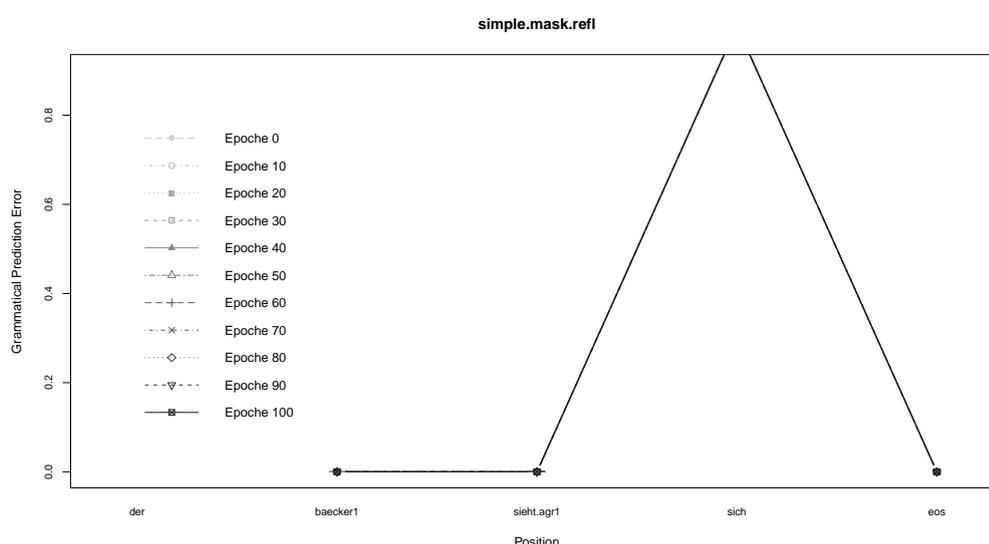


Abbildung 6.44.: Simulation 3b. Verlauf und Entwicklung des globalen Fehlers der Referenzzuweisungs-Komponente, gruppiert nach Epoche: Einfacher Hauptsatz mit Reflexivpronomen.

Dabei hat sich für die komplexeren Satztypen ein interessanter Effekt gezeigt: Für die Aktivierung des Reflexivpronomens nach einem eingebetteten Relativsatz ergibt sich eine verstärkte Aktivierung derjenigen Referenten, die hinsichtlich des Merkmals der Prädikat-Kongruenz mit dem korrekten Antezedens übereinstimmen. Während sich in einfachen Hauptsätzen also keine Präferenzen für eine bestimmte Klasse von Referenten entwickeln, zeigt sich für komplexe Sätze bereits früh (ab Epoche 20) eine vermehrte Aktivierung der Referenten mit übereinstimmender Kongruenz. Dieser Effekt legt nahe, dass es im Verlauf der Verarbeitung über längere Distanzen Einflüsse der grammatischen Informationen auf die Auflösung der Anaphern gibt. Die Interferenzen könnten allerdings gleichermaßen

<sup>53</sup>Die leichte Aktivierung der beiden ausgelassenen Referenten nach Epoche 10 kann vernachlässigt werden. Sie entspricht der anfänglichen allgemeinen Fehlaktivierung, die auch in den vorausgegangenen Simulationen schon beobachtet werden konnte. Fehler und Aktivierungsdetails für die Satzbedingung mit ausgelassener Referentin zeigen identische Verhältnisse.

durch andere Faktoren bedingt sein: Sie könnten durch die intervenierenden Elemente des Kontexts zustande kommen, etwa durch die Aufrechterhaltung und Reaktivierung der Informationen zur Prädikat-Kongruenz für das Verb des Matrixsatzes. Auch die Reduzierung der Trennschärfe der aufrechterhaltenen Kongruenz-Information des Matrixsatz-Subjekts über längere Distanzen, durch die die Informationen nicht mehr nur vom Verb, sondern allgemein genutzt werden, könnte dafür verantwortlich sein. Welcher Natur dieser Effekt auch ist: Wir können auf jeden Fall feststellen, dass die erforderliche Abstraktion zur Bindung von Referenten außerhalb des Trainings-Spektrums in dieser Simulation nicht erworben wurde.

### 6.5.2.2. Weitere Lösungsansätze

Die erste Studie beinhaltet drei Schwächen oder Probleme der verwendeten Stimuli und Trainingsmethodik sowie ein weiteres Problem der verwendeten Architektur, die zum Versagen bei der Aufgabe geführt haben könnten:

1. Zu geringe Anzahl unterscheidbarer Referenten – mangelhafte Generalisierung durch zu geringen Zwang zur Generalisierung.
2. Zu geringe Anzahl ausgelassener Referenten – systematischer Erwerb der Bedeutsamkeit der Nicht-Zuweisbarkeit der ausgelassenen Referenten zu Reflexivpronomen.
3. Vollständiges Feedback für die Referenzzuweisung – vollständiges Feedback für alle anderen Referenten führt zu schlechterer Abstrahierung der Referenzzuweisung.
4. Zu groß gewähltes Hidden-Layer der Referenzzuweisungs-Komponente – Overfitting: Memorierung statt Generalisierung verhindert die benötigte Abstrahierung.

Alle Annahmen bezüglich der Eigenschaften der Stimuli bzw. ihre Pendanten in den neu korrigierten Simulationen sind plausible Aspekte des Erstspracherwerbs: Kinder erwerben eine Vielzahl von Referenten und zugehörigen konkreten Bezeichnungen dieser Referenten. Nur ein geringer Anteil davon wird häufiger mit pronominalen Formen bezeichnet, und nur für wenige Objekte ist eine deiktische Unterstützung (z.B. Zeigen oder Blick auf den aktuellen Referenten der Proform) der Referenzsituation gegeben. Wir haben die vier aufgeführten Aspekte in jeweils eigenen Simulationsreihen untersucht. Wir werden im Folgenden kurz näher erläutern, (a) worin dabei jeweils die vermutete Problematik liegt, (b) wie sie sich auswirkt, (c) welchen Ansatz wir gewählt haben, um das Problem zu überwinden, (d) wie wir den Ansatz umgesetzt haben, und (e) welche Ergebnisse damit erzielt wurden. Bereits vorab sei gesagt, dass wiederum keiner der hier gewählten Ansätze zum gewünschten Resultat geführt hat.

**Anzahl unterscheidbarer Referenten** Die erste Vermutung, die wir untersucht haben, betrifft die Anzahl der Referenten: Eine zu geringe Anzahl unterscheidbarer Referenten könnte eine Abstrahierung der folgenden Art verhindern: *Wenn X zur Klasse Y gehört und im aktuellen Satz die syntaktische Funktion Z erfüllt, so ist X das korrekte Antezedens des Reflexivpronomens (unabhängig davon, ob die konkrete Zuweisung schon gesehen wurde).* Die Annahme einer unterbundenen Abstrahierung kommt deshalb in Betracht, da bei einer kleinen Anzahl von Referenten jeder Referent als Vertreter einer eigenen Klasse

mit eigenen Regelmäßigkeiten gelernt werden könnte. Sollte dies der Fall sein, so genügt es möglicherweise, die Anzahl der Referenten in ausreichendem Maße zu erhöhen. Dies sollte zu einem stärkeren Druck zur Generalisierung und in der Folge zu einer besseren Generalisierungsleistung führen.

Da nicht vorhersagbar ist, was ein ausreichendes Maß ist, haben wir als Schätzwert die Anzahl der Referenten verdoppelt: Unser Lexikon enthält also 16 unterscheidbare Referenten (zu gleichen Anteilen nach Genus und Prädikat-Kongruenz unterschieden), die auf allen syntaktischen Positionen eines belebten Agenten auftreten können. Wiederum wurden zwei Referenten in der Trainingsphase von der Rolle als Antezedens eines Reflexivpronomens ausgeschlossen. Leider brachte dieser Versuch nicht das erwartete Ergebnis. Insgesamt nahm die Performanz der Referenzzuweisung leicht ab, dennoch wurde nach wie vor für die regulären Referenten eine korrekte Zuweisung der Anaphern gezeigt. Für die ausgelassenen Referenten in der syntaktischen Funktion des korrekten Antezedens (das Matrixsatz-Subjekt) wurde jedoch auf der Position des Reflexivpronomens keine Aktivierung gemessen.

**Anzahl ausgelassener Referenten** Unsere zweite Vermutung betrifft die Menge der ausgelassenen Referenten. Eine zu geringe Anzahl ausbleibender Referenten könnte zu dem Effekt führen, dass das fehlende Auftreten eines Referenten mit einem Reflexivpronomen als systematische Ausnahme erworben wird. Die ausgenommenen Referenten bilden also eine eigene Klasse irregulärer Referenten, die nicht mit Reflexivpronomen verwendet werden können (in der Tat entspricht dies den Festlegungen unserer zugrundeliegenden PCFG und ist so gesehen gar kein Erwerbsfehler). Erhöhen wir die Anzahl von Referenten, die nicht als Antezedens des Reflexivpronomens auftreten, verhindert dies möglicherweise diese falsche Generalisierung über einzelne Referenten zugunsten einer abstrakten Generalisierung der Bindung des Reflexivpronomens an die syntaktische Funktion des Satz-Subjekts.

Wiederum kann schwerlich vorab festgestellt werden, wie groß der Anteil ausgenommener Referenten sein muss, damit dieser Ansatz funktioniert. In unserer Umsetzung mit der oben erläuterten Variante des Korpus mit 16 Referenten haben wir als Annäherung genau die Hälfte der Referenten (in gleichen Anteilen nach den Merkmalen Genus und Prädikat-Kongruenz) von der Rolle als Antezedens eines Reflexivpronomens ausgeschlossen. Das Ergebnis blieb auch für diese Simulationsreihe gleich: Keiner der ausgelassenen Referenten wurde auf der Position des Reflexivpronomens aktiviert.

**Feedback für Referenzzuweisung** Unsere dritte Vermutung betrifft das vollständige Feedback für die Referenzzuweisungs-Komponente. Während das vollständige Feedback für die Wortvorhersage im Sinne des selbstüberwachten Lernens (vgl. Kap. 3.1.2) sinnvoll und gerechtfertigt scheint, ist dies bei der Referenzzuweisung weniger der Fall. Schließlich verweisen wir beim normalen Sprachgebrauch nicht ständig mit non-verbalen Mitteln auf den aktuellen Referenten des Wortes, der mit dem aktuellen Wort gemeint ist – und selbst, wenn z.B. deiktisch die Referenzsituation etabliert wird, erfolgt dies nicht synchron zu dem einen Wort. Wir müssen also davon ausgehen, dass die Rückmeldung über die referenzielle Bindung eines Wortes tendenziell asynchron, unvollständig und unpräzise erfolgt.

Möglicherweise ist aber genau das der Schlüssel zur Abstraktion der Referenzbeziehung für Proformen wie etwa Anaphern. So wäre denkbar, dass das vollständige Feedback in unseren Simulationen für alle nicht ausgeschlossenen Referenten dazu führt, dass die Abstrahierung der Referenzzuweisung für Reflexivpronomen nicht geleistet werden kann.

Für die Simulation haben wir wiederum unser Korpus 1 mit zwei ausgelassenen Referenten verwendet, und hiermit verschiedene Methoden für die Umsetzung einer deiktischen Reduktion getestet. Im ersten Ansatz haben wir die Zuweisung von Referenzzuweisung randomisiert: Bei der Umsetzung des Trainings-Korpus in eine Stimulus-Datei im Vektor-Format für die Simulation wurde für jedes Wort anhand einer Randomisierungsfunktion entschieden, ob es auf dem Referenz-Output eine Target-Aktivierung erhält. Diese Methode erwies sich allerdings als ungeeignet, da hierbei über alle Epochen hinweg immer die gleichen Beispiel-Items keine Rückmeldung erhalten. Folglich brach die Referenzzuweisung bereits bei geringen Anteilen verhinderten Feedbacks zusammen. Als Alternative hierzu haben wir das Fehlersignal des Output-Layers der Referenzzuweisungs-Komponente mit einem Störsignal versehen. Damit bleibt die Störung des Fehlersignals unabhängig von der Identität der Beispiele. Während die Zuweisung des Personalpronomen nur minimal davon beeinträchtigt wurde, nahm für das Reflexivpronomen die Qualität der Performanz proportional zum Anteil des Rauschens ab – allerdings in vertretbarem Maße, d.h. selbst bei einem Anteil des Rauschens von über 50% blieb der Mittelwert des globalen Fehlers RAEn unter 0.5. Allerdings ergab sich dadurch keine Verbesserung der Performanz für die ausgelassenen Referenten, die bei der Zuweisung von Personalpronomen wiederum keine Aktivierung erhielten.

**Kontrolle der Hidden-Layer** Als letzte Option blieb noch die Variation des Hidden-Layers der Referenzzuweisungs-Komponente: Bei der Wahl eines zu großen rekurrenten Layers im SRN erhöht sich die Gefahr des Overfitting (vgl. z.B. S. L. Frank, 2006c): Das Netzwerk memoriert die Beispiele – bzw. die Kollokationen der unterscheidbaren Wörter, so als würde jedes Wort eine eigene syntaktische Klasse bilden – anstatt über sie zu generalisieren und in übergreifende syntaktische Klassen zu kategorisieren. Folglich könnte die Unfähigkeit unserer Netzwerke, abstrakte Generalisierungen für die Bindung von Reflexivpronomen zu bilden, schlichtweg der Effekt eines zu groß gewählten Hidden-Layers sein, während ein kleineres Hidden-Layer das Netzwerk zu einer geeigneteren Generalisierung zwingen würde.

Wir haben wieder unser Korpus 1 mit zwei ausgelassenen Referenten verwendet, und in einer Simulationsreihe stufenweise die Größe des Hidden-Layers der Referenz-Komponente verringert. Während die allgemeine Performanz für die im Trainings-Set vorhandenen Referenten auch für sehr kleine Hidden-Layer (unter 10 Units) noch vertretbar gut blieb, führte die Variation wiederum nicht zu Verbesserungen der Performanz für die ausgelassenen Referenten.

### 6.5.2.3. Diskussion

Unsere Bemühungen haben keinen Erfolg gebracht: Keiner der getesteten Ansätze konnte die Identitätsfunktion für den neuen Referenten auch nur annähernd lösen. Sind SRNs

generell nicht dazu befähigt, diese Aufgabe zu leisten? Zwar ist eine nicht erfolgreiche Simulation noch kein Beweis dafür, dass SRNs entsprechende Aufgaben nicht leisten können, und es ist auch nicht auszuschließen, dass spezifischere Variationen von Architektur, Stimuli oder Trainingsmethoden doch zum Erfolg führen können. Allerdings wurde nach unserer Kenntnis bislang nicht das Gegenteil bewiesen. Und selbst, wenn eine erfolgreiche Simulation geleistet wird, wird das Ergebnis stark von der spezifischen Realisierung der Modellierung abhängig sein – eine Abhängigkeit, für die eine gute Begründung geliefert werden muss. Aus unseren Simulationen müssen wir jedenfalls vorübergehend den Schluss ziehen, dass dieser Systematizitätsaspekt andere kognitive Mechanismen fordert als diejenigen, die mit SRNs simuliert werden können.

### 6.5.3. Korpus 3c: Vergleichsstudie ohne Referenzzuweisung

Welchen Effekt hat die Integration der Referenzzuweisungs-Komponente auf die Performanz und die internen Repräsentationen der Wortvorhersage-Komponente? Von dieser Fragestellung ausgehend haben wir ein einfaches SRN mit gleichen Spezifikationen für Architektur und Trainings-Parameter<sup>54</sup> wie zuvor in Simulationsstudie 2 mit Korpus 2 trainiert und verschiedene Kennwerte der Netzwerke für Epoche 10 mit den Ergebnissen und Merkmalen der Netzwerke aus der ersten Simulationsstudie verglichen. Neben der allgemeinen Qualität der Performanz ist der Verlauf des Fehlers über die verschiedenen Satztypen des Test-Sets von Interesse: Zeigen sich hier auffällige Unterschiede, und wenn ja, worauf sind diese zurückzuführen? Ebenfalls relevant ist die Strukturierung der Aktivierungsmuster des Hidden-Layers – mit besonderem Augenmerk auf die Wortklassen, die in der integrierten Simulation einen Verweis auf einen Referenten enthalten.

#### 6.5.3.1. Performanz und interne Repräsentation

Nach 10 Epochen ergibt sich für die allgemeine Performanz ein gemittelter Fehler von  $\bar{x}GPE_n = 0.130$  (Standardabweichung  $\sigma = 0.092$ ). Für die Verben wird ein Fehlerwert von  $\bar{x}GPE_n = 0.149$  (Standardabweichung  $\sigma = 0.055$ ) erreicht. Dies sind tendenziell bessere Ergebnisse als in Simulation 2 (vgl. Kap. 6.4.3.1) – womit sich andeutet, dass unsere Vermutung eines Bootstrapping-Effekts nicht bestätigt werden kann. Auch im Verlauf der Entwicklung in den frühen Epochen läßt sich darauf kein Hinweis finden. Da in dieser Simulation ausschließlich Informationen über Wortübergangs-Wahrscheinlichkeiten gegeben sind, wäre denkbar, dass diese Veränderung der vorhandenen Informationsstruktur zu einer Verstärkung des LSK-Effekts führt. Allerdings deutet sich im Verlauf des globalen Fehlers für die verschiedenen Satztypen, für die sich keine Unterschiede zu den Fehlerverläufen in Simulation 2 zeigen, bereits an, dass keine entsprechenden Effekte zu erwarten sind. Dies hat sich in den Detail-Analysen der Aktivierungsverteilung bestätigt: Die Aktivierungen erbringen keine Hinweise zu qualitativen oder quantitativen Unterschieden zwischen den Satzbedingungen im Vergleich zu Simulation 2. Der einzige Unterschied, der sich konstant

---

<sup>54</sup>Insgesamt wurden 10 Netzwerke für 20 Epochen mit dem Algorithmus Backpropagation Through Time mit einer Lernrate von  $\epsilon = 0.15$ , einem Momentum von  $momentum = 0.3$  und einer initialen Randomisierung der Gewichte innerhalb von  $initialweightrange = 0.1$  trainiert. Die Größe des Hidden-Layers umfasst 80 Units, Input- und Output-Layer jeweils 50 Units. Der Output-Layer erhielt eine Softmax-Funktion.

in allen Satzstrukturen findet, ist ein etwas höherer Fehler in den frühen Epochen auf dem Relativpronomen. Allerdings lässt sich hierfür in den Aktivierungs-Details keine eindeutige Ursache identifizieren.

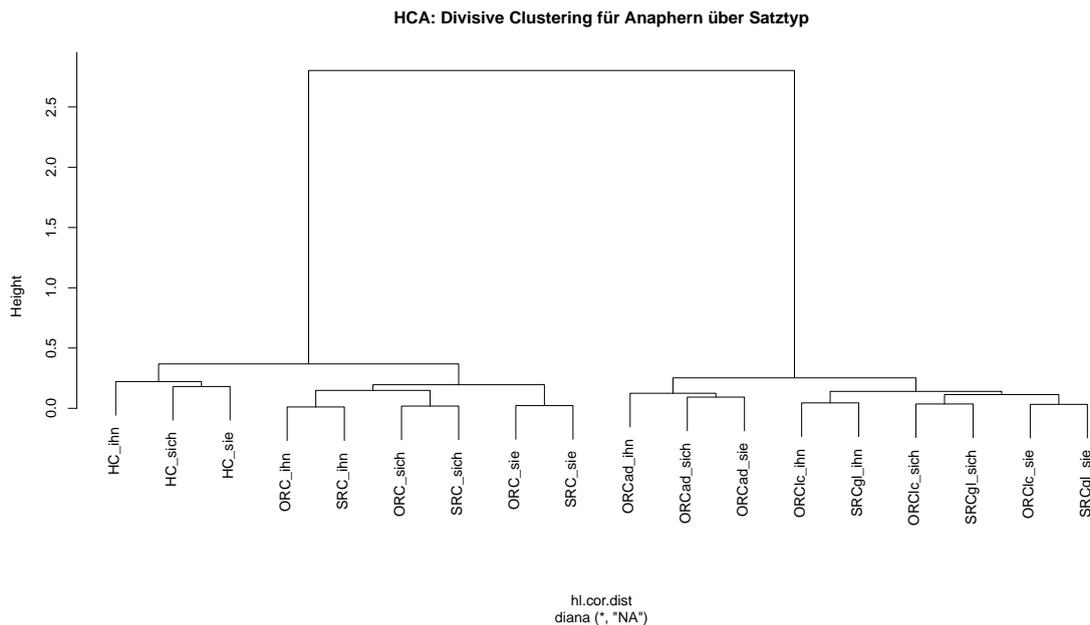


Abbildung 6.45.: Simulation 3c. Hierarchische Cluster-Analyse (divisiv) der Hidden-Layer-Aktivierungen: Anaphern in Abhängigkeit ihres syntaktischen Kontexts.

Etwas interessanter ist die Struktur des Hidden-Layers. Abbildung 6.45 zeigt die Cluster-Struktur für die Anaphern in Abhängigkeit ihres syntaktischen Kontexts. In Simulation 2 fanden sich unterschiedliche interne Strukturierungen des Zustandsraums über die beiden Komponenten, wobei sich für die Wortvorhersage-Komponente eine geeignetere Struktur für die Wortvorhersage entwickelte. Dieser Effekt wird in Simulation 3c noch verstärkt: Während in Simulation 2 die Anaphern zusammen geclustert wurden, die in einen Hauptsatz gegenüber einem Nebensatz eingebettet sind, finden wir hier eine weitere klare Untergliederung der Hauptsatzeinbettung für einfache und komplexe Hauptsätze (mit vs. ohne eingebettetem Relativsatz). Im Unterschied zur Cluster-Struktur in Simulation 2 zeigt sich auch, dass die Hauptunterscheidung zwischen Haupt- und Nebensatzeinbettung mit deutlich größerer Distanz zwischen den Clustern realisiert ist. In einer eFA für einzelne Wortformen (ohne Abb.) zeigt sich zudem, dass die Trennung der Relativpronomen, die in der Simulation 2 entlang des Merkmals der Prädikat-Kongruenz erfolgte, hier aufgehoben ist.

### 6.5.3.2. Diskussion

In Anbetracht der Ergebnisse der reduzierten Simulation 3c können wir nicht darauf schließen, dass in den anderen Simulationen der Bootstrapping-Effekt für den Erwerb syntaktischen Wortklassen erfolgt ist, den wir hinsichtlich der zusätzlichen Strukturierungsleistungen für die Referenzzuweisungs-Komponente erwartet hätten. Auch die Details der

Performanz geben keine Hinweise auf einen Unterschied der reduzierten Architektur. Die detailreichere interne Strukturierung in der vollständigen Architektur scheint für die Wortvorhersage keine Vorteile zu bringen. Generell zeigen die internen Repräsentationen eine weitere Spezialisierung der Struktur des Hidden-Layers für die Wortvorhersage-Aufgabe. Die Tatsache, dass R. Frank et al. (2005) eine solche reduzierte Repräsentation als Grundlage für die Referenzzuweisungs-Aufgabe verwenden, dürfte einige negative Aspekte ihrer Ergebnisse erklären. Die internen Strukturen hingegen, die wir für Simulation 2 festgestellt haben, scheinen demgegenüber eine integrative Lösung darzustellen, die zwischen den beiden Komponenten für die Bewältigung der verschiedenen Aufgaben ausgehandelt wurde.

### 6.6. Diskussion des Modells AnaSRN

Das Modell AnaSRN hat sich als geeignet für die Simulation der Auflösung anaphorischer Referenzen erwiesen. Durch die Erweiterung der einfachen SRN-Architektur und die dadurch ermöglichte Integration weiterer Aufgaben können wir Gegenstände der Sprachverarbeitung modellieren, die mit ausschließlicher Wortvorhersage nicht möglich wären. Damit werden weitere Einblicke in die Eigenschaften dieser Architektur eröffnet. Zudem erlaubt unsere Architektur neben der Modellierung psycholinguistischer Daten auch die Vorhersage von Effekten in der referenziellen Sprachverarbeitung. Mit dem Modell wurde der Erwerb und die Verarbeitung von Anaphern nach strukturellen Bindungs-Constraints erfolgreich simuliert. Dabei erzielt es auch in komplexen syntaktischen Kontexten eine sehr gute Performanz. Grundsätzlich erhalten korrekte Antezedenzen gegenüber unzulässigen nach wenigen Epochen eine wesentlich stärkere Aktivierung. Falsche, übermäßige und fehlende Aktivierungen können als Effekte im Sinne der Fehler-Korrespondenz des Modells (vgl. Kap. 2.1) interpretiert werden. Hierfür haben wir in verschiedenen Simulationen verschiedene Aspekte der Stimuli variiert, um die Verarbeitung unter verschiedenen Bedingungen zu untersuchen und spezifische Verarbeitungseigenschaften zu identifizieren.

#### 6.6.1. Ableitung empirischer Hypothesen

Welche empirischen Vorhersagen können wir aus den Simulationen ableiten? Für die syntaktische Verarbeitung, die über die Wortvorhersage simuliert wurde, konnten wir verschiedene Ursachen für Interferenz-Effekte identifizieren, die von Lokalität und Distanz über Frequenz und Regularität bis zu LSK reichen. Diese Effekte wurden bereits in früheren Studien beschrieben und auf Aspekte der Satzverarbeitung unter anderem im Sinne von Komplexität, Parallelverarbeitung, Erfahrung, Antizipation und interindividuellen Differenzen bezogen, die wir ausführlich in Kapitel 3.2 diskutiert haben. Mit AnaSRN haben wir diese Aspekte auf den neuen Gegenstand anaphorischer Referenzen übertragen. Relevant werden dabei vor allem die Ergebnisse der Referenzzuweisung. Wir fassen kurz zentrale Befunde der Simulationen zusammen, die wir als empirisch überprüfbare Vorhersagen des Modells erachten.

Für die Referenzzuweisung können wir zunächst eine allgemeine Auswirkung der Komplexität der Satzstruktur auf die Anaphernauflösung feststellen. Je komplexer die Satzstruk-

tur (etwa durch Einbettung von Relativsätzen) oder je seltener die Satzstruktur (etwa durch vorangestellte Nebensätze) ist, in der die Anapher erscheint, desto stärker wird die Präzision ihrer Auflösung beeinträchtigt. Dabei können sich kontextunabhängige Interferenzen sowie in spezifischen Satzstrukturen Interferenzen des syntaktischen Kontexts auf die Auswahl der Koreferenz-Kandidaten auswirken. Hier nehmen alternative Referenten dann Einfluss auf die Verarbeitung, wenn sie in ihren grammatischen Merkmalen (etwa Genus oder Prädikat-Kongruenz) mit dem korrekten Antezedens übereinstimmen (in Entsprechung zu Befunden u.a. von Badecker & Straub, 2002; Sturt, 2003). Für Personalpronomen konnten wir zeigen, dass deren Verarbeitung durch LSK-Effekte beeinflusst wird: Distraktoren der lokalen Subsequenz werden geringfügiger berücksichtigt, obwohl sie zulässige Antezedenzen sind, während der Referent des unzulässigen Matrixsatz-Subjekts verstärkt aktiviert wird. Damit wird eine Erweiterung des Wirkungskreises für das Phänomen der lokalen syntaktischen Kohärenzen (Tabor et al., 2004; Konieczny, Müller, Hachmann et al., 2009) auf die Anapherauflösung vorhergesagt. Reflexivpronomen werden geringfügiger durch den syntaktischen Kontext und stärker durch kontextunabhängige Interferenzen geprägt.<sup>55</sup> LSK-Effekte treten vor allem dann auf, wenn das Netzwerk seine Interpretation verstärkt auf Wortabfolge-Regelmäßigkeiten stützen muss, etwa weil die satzinterne Interpunktion fehlt.

Formulieren wir daraus einige allgemeiner gefasste Hypothesen für empirische Studien: In einfachen Sätzen ist die Anapherauflösung präziser oder erfolgt schneller als in komplexen Sätzen.<sup>56</sup> Interferenz-Effekte sind vor allem durch alternative Referenten zu erwarten, die in Genus und Numerus mit dem korrekten Antezedens übereinstimmen. Für LSK werden Einflüsse auf die Anapherauflösung erwartet. Diese sollten sich auf Personalpronomen und Reflexivpronomen unterschiedlich auswirken.<sup>57</sup> Zudem sollten sie sich stärker auswirken, wenn Informationen über die Satz-Struktur bzw. -Gliederung (etwa Satzzeichen, Intonation, Pausen) reduziert werden. Allgemein geht mit diesen Effekten eine Verletzung der Annahme der globalen Konsistenz der Interpretation einher. Diese Hypothesen sind ein zentraler Ausgangspunkt für das empirische Projekt AnaLoCo, das im folgenden Kapitel 7 vorgestellt wird.

Die hier vorgestellten Simulationsstudien zeigen nur einen kleinen Ausschnitt möglicher Satzstrukturen und anaphorischer Formen, die mit dem Modell untersucht werden könnten. Interessant wäre die Implementierung einer umfangreicheren Grammatik (mit de-

<sup>55</sup>Solche Unterschiede der bedingenden Constraints für verschiedene anaphorische Formen werden auch in der Hypothese formenspezifischer multipler Constraints (Brown-Schmidt et al., 2005; Kaiser et al., 2009) angenommen.

<sup>56</sup>Die Linking-Hypothese hierfür ist, dass der Fehleranteil in der Aktivierungsverteilung der Referenzzuweisungs-Komponente die Einfachheit und Eindeutigkeit der Koreferenz-Bestimmung sowie ihre Anfälligkeit für multiple Interferenz-Faktoren widerspiegelt. Je präziser und fehlerfreier die Auswahl im SRN erfolgt, desto weniger Interferenz-Effekte sind bei der natürlichen Sprachverarbeitung zu erwarten, und desto geringer sollte der kognitive Aufwand bei der Auflösung und beim Ausschluss alternativer Koreferenten sein. Dies sollte sich in entsprechenden Experimental-Paradigmen in geringeren Fehlerraten, kürzeren Latenzen und präziserer Auswahl (z.B. bei der Aufmerksamkeitslenkung in Eyetracking-Experimenten) niederschlagen.

<sup>57</sup>Gerade in diesem Aspekt der Hypothese sind wir uns nicht schlüssig, wie viel Gewicht den Simulationsergebnissen für die Hypothesenbildung beigemessen werden sollte. Anderen Befunden folgend nehmen wir eher an, dass gerade Reflexivpronomen durch LSK-Effekte beeinflusst werden sollten, da ihre Auflösung stärker durch strukturelle Constraints bestimmt wird.

taillierteren semantischen Constraints), die die Modellierung der Daten aus verschiedenen empirischen Studien zur Anaphernaufflösung (etwa zu den Stimuli der Experimente von Badecker & Straub, 2002; Kennison, 2003; Sturt, 2003; Runner et al., 2006; Kaiser et al., 2009, vgl. Kap. 5.3.2) innerhalb einer Simulation ermöglicht. In diesem Zusammenhang mit möglichen Erweiterungen ist die linguistische Abschlussarbeit von Niefind (2009) zu erwähnen, in der das Projekt AnaSRN weiterverfolgt wurde<sup>58</sup>: In dieser wird eine Grammatik untersucht, die sich stärker an Gegebenheiten natürlicher Sprache orientiert. Dies betrifft unter anderem den zugrundeliegenden theoretischen Ansatz zur Anaphernaufflösung, für den ein Wechsel von der strukturell-syntaktisch ausgerichteten Bindungstheorie zur funktional ausgerichteten *Perspective Hypothesis* von MacWhinney (2008) erfolgte. Im Zentrum der Untersuchung steht die Unterscheidung von Anaphern und Alternativkonstruktionen mit Demonstrativpronomen (*dieser, diese*), über die eine Untersuchung der Anbindungspräferenzen (Subjekt- vs. Objekt-Anbindung) erfolgt. Proformen treten dabei nur für Referenten auf, die tatsächlich bereits in demselben Satz eingeführt wurden. Verglichen werden Sätze, in denen die Anaphern hinsichtlich ihrer Position im Satz und hinsichtlich ihres Antezedens variieren. Die allgemeine Performanz des Modells erbrachte gute Ergebnisse für die erforderlichen Referenzzuweisungen. Untersuchungen zu weitergehenden Hypothesen, nach denen ein Bootstrapping-Mechanismus für die bessere Erfüllung syntaktischer Constraints durch die referenziellen Constraints oder eine trainingsunabhängige Systematizität erwartet wurde, erbrachten aber keine Ergebnisse.

### 6.6.2. Kritische Evaluation des Modells

Wir halten das Modell AnaSRN zwar für ein geeignetes Modell zur Simulation der Anaphernaufflösung, doch in verschiedener Hinsicht ist auch Kritik an ihm zu üben. Zunächst einmal können die zugrundeliegenden SRNs unserer Einschätzung nach nicht das Problem der Identitätsfunktion lösen, was die Anwendbarkeit dieser Architektur für manche Fragestellungen einschränkt. Im Gegenzug bieten sie durch ihre inhärenten Eigenschaften und Beschränkungen eine interessante Arbeitshypothese für die Simulation menschlicher kognitiver Prozesse und einen guten Ansatz für die Identifikation und Erklärung von deren Eigenschaften und Beschränkungen. Unseres Erachtens sollten SRNs ohnehin nicht als umfassende Erklärungsmodelle der menschlichen kognitiven Kapazität und Leistungsfähigkeit verstanden werden. Sie bieten vielmehr einen Ansatz für einen Ausschnitt der relevanten Teilaspekte der Verarbeitung, unter anderem Sequenzverarbeitung, induktive Klassifizierung und Constraint-Satisfaction-Prozesse. Wir müssen SRNs von vorne herein als eine beschränkte Architektur verstehen, die nur für diese Teilaspekte relevante Aussagen generieren kann. Aus dieser Perspektive sind sie definitiv nützlich für die gezielte Simulation und Vorhersage spezifischer psycholinguistischer Daten.

Eine weitere Einschränkung des Modells AnaSRN ist, dass es wie SRNs allgemein nur die Simulation von Effekten ermöglicht, die online erfolgen, nicht aber weitergehende Effekte und Mechanismen, die über den Augenblick der ersten Verarbeitung hinausgehen. In dieser Hinsicht könnte unser Modell etwa mit der Early-Filter-Hypothese (Nicol & Swin-

---

<sup>58</sup>Das Modell wurde im Rahmen eines Projektseminars am Institut für Informatik und Gesellschaft (Universität Freiburg) diskutiert. In einer Projektgruppe unter Leitung des Autors wurden weitere Simulationen durchgeführt. Aus diesem Projekt heraus entstand die Arbeit von Niefind (2009).

ney, 1989) oder mit dem zweistufigen Ansatz von Sturt (2003) verglichen werden. Eher als die Interpretation der Anaphern leistet es die Aufgabe eines initialen Filters, der aufgrund morphosyntaktischer, syntaktischer und semantischer Constraints eine Auswahl der in Betracht zu ziehenden Referenten trifft. Denn alleine aus der Antizipation der möglichen Folgewörter und der Zuweisung möglicher Referenten auf einer aktuellen Position ergibt sich noch nicht die tatsächliche Interpretation im Sinne der Auswahl des Wortes bzw. bei der Anaphernauflösung die Bestimmung des Antezedens. Des Weiteren kann es einerseits sein, dass unsere Verarbeitung online tatsächlich nur gut genug oder in lokalen Fenstern erfolgt. Andererseits können andere Prozesse über den Augenblick hinaus zu einem globalen Verstehen eines Satzes führen, für das temporäre Interpretationsfehler bereinigt werden und die global korrekte Interpretation identifiziert wird. Solche Prozesse können mit SRNs ohne Erweiterung um Verstehenskomponenten und Bedeutungsstrukturen nicht erfasst werden. Einfache Erweiterungen des Modells könnten Aspekte der Interpretation von Anaphern ermöglichen: Denkbar wäre etwa die Ergänzung um einen zusätzlichen Input, der dem Netzwerk ein Diskurs-Modell als Kontext der Verarbeitung bietet, für das im Satz textuelle Kohärenz hergestellt werden muss (vergleichbar etwa mit CIA-Net, Mayberry et al., 2009). Die hier zu leistenden Inferenzprozesse genügen möglicherweise, um den Auswahlprozess des als Antezedens interpretierten Referenten zu leisten. Zudem wäre dies ein Ansatz, um die Systematizität des Modells zu erweitern (im Sinne einer semantisch gestützten Systematizität, vgl. Hadley et al., 2001). Ein solches Vorgehen der Integration von Sprache und Gesamtsituation verfolgen wir im Projekt SynSem, das in Kapitel 10 vorgestellt wird.

Konkreter auf die vorgestellten Simulationen bezogen stellen sich für AnaSRN auch einige Fragen, die die Interpretation der Performanz des Modells betreffen. Nicht ganz unproblematisch ist die lange erforderliche Lernzeit der Referenzzuweisungs-Komponente für die pronominalen Formen. Einerseits muss festgehalten werden, dass in allen Simulationen die Satzbedingungen für beide pronominalen Formen bereits ab Epoche 10 (bzw. in einigen Fällen ab Epoche 20) die korrekten Antezedenzen die stärkste Aktivierung erhalten, andererseits zeigen sich noch längerfristig Verbesserungen der Zuweisungs-genauigkeit. Eine offensichtliche Erklärung, warum die Lösung bzw. Verbesserung dieser Aufgabe so viel länger dauert, gibt es nicht. Eine Möglichkeit könnte sein, dass dies ein Artefakt der Vermittlung der Aufgabe über die Wortvorhersage-Komponente ist. Insofern wäre in Frage zu stellen, ob diese Vermittlung von Referenzinformationen über syntaktisch orientierte Repräsentationen eine psycholinguistisch valide Annahme ist. Allerdings zeigt sich mit der unterschiedlichen Lerngeschwindigkeit für Referenz und Wortvorhersage eine gewisse Überschneidung zum kindlichen Spracherwerb, bei dem die initiale Verwendung pronominaler Formen vergleichsweise spät erfolgt.

In diesem Zusammenhang stellt sich auch die Frage der Auswahl der zu interpretierenden Epoche: Welche Epoche soll als psycholinguistisch relevant interpretiert werden? Sowohl für die Wortvorhersage als auch für die Referenzzuweisung könnte man als Untergrenze für die Verwendbarkeit einer Epoche eine Performanzleistung erwarten, bei der der Vorhersagefehler einen gewissen Schwellenwert unterschreitet, da die gefundenen Effekte ansonsten möglicherweise auf dem unausgereiften Zustand des Netzwerks beruhen – wie etwa von Konieczny, Ruh und Müller (2009) an der Studie von MacDonald und Christiansen (2002) kritisiert wurde. Aber wie soll dieser Schwellenwert festgelegt wer-

den? Eine Möglichkeit wäre, dass die korrekten Wortformen stärker aktiviert sein müssen als die falschen Wortformen. Dabei müssten aber unter Umständen konsistente falsche Vorhersagen unberücksichtigt bleiben. Eine weitere Möglichkeit wäre, die überwiegend korrekte Aktivierung spezieller Wortformen vorauszusetzen: So stellt etwa die Vorhersage der korrekten Kongruenz der Verben ein Kriterium dar, an dem die erfolgreiche Auflösung von Fernabhängigkeiten überprüft werden kann. Allerdings bleibt auch hier wieder eine Frage offen: Genügt es bereits, wenn die Mindestanforderung in einfachen Sätzen erfüllt wird? Oder müssen sie auch in komplexen Sätzen erfüllt werden? Wenn ja, in welchen? Alternativ könnte man ausschließlich die Epoche zur Analyse heranziehen, ab der sich im Training keine Entwicklung der Fehlerverbesserung mehr zeigt. Allerdings nähert man sich hier auch einer Schwelle, ab der sich möglicherweise vermehrt Übergeneralisierungen finden. Zudem wäre in Frage zu stellen, warum ausgerechnet dieser Zustand die beste Vergleichbarkeit mit empirischen Daten eröffnen sollte. Aus Sicht der Spracherwerbsforschung wäre zu fragen, welche Epochenanzahl hierbei legitim und welche eher unrealistisch ist.<sup>59</sup> Nicht zuletzt werden mit der ausschließlichen Betrachtung der letzten Epoche Entwicklungsverläufe und Erfahrungsunterschiede aus der Interpretation ausgeschlossen, die zum Beispiel zur Erklärung interindividueller Unterschiede herangezogen werden können (vgl. Wells et al., 2009). Wenn eine festgelegte Mindestleistungsgrenze erreicht wurde, besteht immer noch die Frage, aus welchen Epochen wir anschließend Effekte berücksichtigen dürfen. Können unabhängige Fehler verschiedener Epochen als Effekte interpretiert und mit empirischen Daten verglichen werden? Oder sollte ein Fehler über mehrere Epochen hinweg Bestand haben, damit er interpretierbar wird (wobei die Größe der Epochen allerdings auch wieder willkürlich bestimmt werden kann)?

Die Interpretation der Ergebnisse konnektionistischer Simulationen und ihr Vergleich mit empirischen Daten bleibt nach wie vor eine Herausforderung der konnektionistischen Psycholinguistik. Dies betrifft zum einen die Wahl der Fehlermaße und sonstiger Daten, die zum Vergleich herangezogen werden, und zum anderen die Linking-Hypothese zwischen simulierten und empirischen Daten: Welche Daten der Simulation können mit welchen Daten der empirischen Forschung verglichen werden? Ist die globale Fehlerrate des Netzwerk-Outputs ein guter Vergleich für Ratings, Lesezeiten und Blickbewegungen? Und nicht zuletzt betrifft dies die oben diskutierte Wahl des Performanzausschnitts des Netzwerks mit der Auswahl der als relevant erachteten Epochen. Eine Diskussion der Kriterien, die für diese Fragen anzulegen sind, könnte die Vergleichbarkeit und Interpretation von Ergebnissen der konnektionistischen Psycholinguistik wesentlich verbessern. Zusätzlich muss diese Diskussion aber auch für jede Simulation spezifisch geführt werden, um die jeweiligen individuellen Details und Eigenschaften zu berücksichtigen – in der Studie AnaSRN haben wir das Ziel verfolgt, diesem Anspruch gerecht zu werden.

---

<sup>59</sup>Es dürfte allerdings schwierig werden, in Anbetracht der erheblichen Unterschiede zwischen den verwendeten künstlichen Grammatiken der Simulationen und natürlichen Sprachen (u.a. viel größere Menge sprachlichen Inputs, viel komplexere Grammatik, viel mehr Informationen anderer sprachlicher sowie nicht-sprachlicher Aspekte) einen sinnvollen Vergleich zu finden.

## 7. AnaLoCo: Anaphern und lokale Kohärenzen

Im folgenden Kapitel werden die im Rahmen eines größeren empirischen Projekts durchgeführten Experimente vorgestellt. Wir haben dem Projekt den Arbeitstitel *AnaLoCo* gegeben, eine Kurzform für Anaphora resolution in Locally syntactic Coherences. Die Idee zum Projekt hat sich parallel zum Modellierungsprojekt *AnaSRN* entwickelt. Einige der Fragen und Hypothesen, die das Modellierungsprojekt angetrieben haben, liegen dem empirischen Projekt zugrunde. Damit schließt sich auch der Kreis klassischer kognitionswissenschaftlicher Forschung: Empirisch-experimentelle Methoden und synthetische Modellierungsmethoden greifen ineinander, informieren sich gegenseitig und führen gemeinsam zu neuen Perspektiven auf den Untersuchungsgegenstand.

Ziel ist die Untersuchung der Satzverarbeitung bzw. des Satzverstehens unter dem Gesichtspunkt des Phänomens lokaler syntaktischer Kohärenzen (im Folgenden LSK) und ihrer Auswirkungen auf die Online-Verarbeitung von Anaphern. Gegenstand sind hier Personal- und Reflexivpronomen, die innerhalb einer LSK auftreten. In der Untersuchung geht es speziell um die Frage, ob LSK einen (zeitlich beschränkten) Einfluss darauf haben, welches Antezedens für eine Anapher als Koreferent in Betracht gezogen wird. In einer Serie von drei Experimenten – einer computergestützten Lesezeitmessung mit Self-Paced Reading, zwei Blickbewegungsmessungen im Visual-World-Paradigma sowie einer Normierungsstudie zum Rating der verwendeten Materialien in der ersten Eyetracking-Studie – gehen wir dieser Frage auf den Grund (vgl. auch Weldle, Konieczny, Wolfer et al., 2010; Konieczny, Weldle, Wolfer et al., 2010). Die gefundenen Effekte haben Konsequenzen sowohl für Annahmen zur Verarbeitung von LSK als auch für theoretische Aspekte der Online-Auflösung von Anaphern.

### 7.1. Temporäre lokale Bindungsdomänen

Unseren Studien liegt folgende Ausgangsfrage zu Grunde: Beeinflusst die lokale syntaktische Struktur die Online-Auflösung anaphorischer Referenzen? Das heißt, etwas konkreter formuliert: Wird die Verarbeitung von Anaphern genauso von lokalen Fenstern der Strukturanalyse geleitet oder zumindest beeinflusst, wie es frühere Untersuchungen für verschiedene syntaktische Strukturen nahelegen? In diesen Sätzen wurde durch eingebettete LSK die Beurteilung der syntaktischen Anbindung hinzukommender Elemente sowie der thematischen Rolle dieser Elemente beeinflusst. Wir beschäftigen uns in der folgenden Studie also mit der weitergehenden Frage, ob LSK auch die Auswahl der in Betracht gezogenen Antezedenzen für anaphorische Referenzen und damit deren Interpretation beeinflussen, wenn die Anaphern innerhalb der LSK liegen.

### 7.1.1. Gegenstand und Fragestellung

Dies wollen wir an einem Satz exemplifizieren, der der Einfachheit halber in englischer Sprache verfasst ist: Aufgrund des Prinzips A kann *himself* in Beispiel 25a nur mit *Ken* koreferieren. Genau dasselbe gilt auch für Beispiel 25b. Allerdings enthält dieser Satz eine LSK – die syntaktisch kohärente Sequenz *John<sub>j</sub> sees himself<sub>j</sub>* – innerhalb der *himself* isoliert vom globalen Kontext mit *John* koreferieren könnte. Parallel dazu verhält sich die Situation für Personalpronomen wie folgt: Aufgrund des Prinzips B darf im Beispiel 25c *him* nicht mit *Ken* koreferieren, ansonsten aber mit jedem beliebigen anderen grammatisch kompatiblen Referenten. Genau dasselbe gilt für Beispiel 25d, in dem der Satz aber wiederum eine LSK enthält, *John<sub>j</sub> sees him<sub>i/\*j/k</sub>*. Isoliert vom globalen Kontext könnte innerhalb der LSK *him* nun auch mit *Ken* koreferieren, aber nicht mehr mit *John*.<sup>1</sup>

- (25)
- a. Ken<sub>i</sub> sees himself<sub>i</sub> in the mirror.
  - b. Ken<sub>i</sub> who likes John<sub>j</sub> sees himself<sub>i/\*j</sub> in the mirror.
  - c. Ken<sub>i</sub> sees him<sub>\*i/k</sub> on the picture.
  - d. Ken<sub>i</sub> who likes John<sub>j</sub> sees him<sub>\*i/j/k</sub> on the picture.

In unserer Forschungsfrage konvergieren die in Kapitel 5 behandelten empirischen Untersuchungsdomänen: Das syntaktische Phänomen kohärenter Subsequenzen innerhalb von Sätzen wird in seiner Auswirkung auf eine bislang in diesem Bereich nicht beachtete Untersuchungsdomäne angewandt – die Interpretation anaphorischer Referenz.

Die empirische Untersuchung der Online-Verarbeitung anaphorischer Referenzen beschäftigt sich vor allem mit der Frage, durch welche Faktoren bestimmt bzw. beeinflusst wird, welches im Satz vorausgegangene Agens als Antezedens in Betracht gezogen und schließlich als korrektes Antezedens interpretiert wird. Dies ist gerade dann interessant, wenn innerhalb eines Satzes (bzw. in einem größeren Diskurs-Kontext über mehrere Sätze hinweg) mehrere Agentia eingeführt wurden, die unter semantischen und morphosyntaktischen Gesichtspunkten als Koreferent in Betracht kommen. Auf unsere Fragestellung übertragen soll untersucht werden: Inwiefern beeinflussen LSK die Filterung und Auswahl des Agens bei der Verarbeitung einer Anapher innerhalb dieser LSK, wenn im Satz mehrere Agentia eingeführt wurden, die hinsichtlich der Übereinstimmung ihres Genus, ihres Numerus und ihrer semantischen Plausibilität als Agens oder Patiens der Handlung als mögliche Antezedenzen für die koreferenzielle Auflösung der Anapher in Betracht kommen.

In unseren Studien finden sich positive Befunde, die in mehrerlei Hinsicht Konsequenzen für die Diskussionen in den beiden Untersuchungsdomänen haben: Erstens hinsichtlich in Betracht zu ziehender Faktoren der Online-Anaphernaufflösung, und damit auch hinsichtlich der Annahmen zur Rolle und Wirkungsweise struktureller Faktoren. Zweitens hinsichtlich des zeitlichen Verlaufs der Anaphernaufflösung und der Tiefe ihrer unmittelbaren Verarbeitung. Drittens für die theoretische Ansätze zur Verarbeitung lokaler syntaktischer Kohärenzen, deren zugrundeliegenden Prozesse sich in der Folge auch auf die Verarbeitung referenzieller und koreferenzieller Bindung auswirken.

---

<sup>1</sup>Indizes markieren entsprechend der gängigen Methodik die Koreferenz nominaler und pronominaler Elemente, bei gleichem Index verweisen diese also auf den gleichen Referenten. Ein Asterisk vor dem Index markiert eine grammatikalisch unzulässige Koreferenz-Beziehung.

**Lokale strukturelle Constraints der Anaphernaufflösung** Der lokale syntaktische Kontext kann als weiterer Faktor identifiziert werden, der die Bindung anaphorischer Referenten beeinflusst. Dies zeigt, dass sich strukturelle Constraints in der Online-Verarbeitung auf verschiedenen Ebenen (eben nicht nur auf der globalen Ebene des Satzes) auswirken.

**Unmittelbarkeit der Verarbeitung von Anaphern** Dies kann als Hinweis auf eine schnellere bzw. frühzeitig tiefe Verarbeitung von Anaphern (bis zur tatsächlichen Zuweisung des Koreferenten) interpretiert werden. Folglich müssen LSK als neues Untersuchungsparadigma berücksichtigt werden, um weitere Strukturen bzw. linguistische Domänen zu identifizieren, in denen unmittelbar lokale Verarbeitungsprozesse wirken.

**Verarbeitungstiefe lokaler syntaktischer Kohärenzen** Für die postulierten Effekte der Verarbeitung von LSK legt dies nahe, dass die lokale Verarbeitung tiefer greift als bislang angenommen – in diesem Fall bis in die Ebene referenzieller Interpretation. Dies kann bedeuten, dass die für die LSK-Effekte verantwortlichen Verarbeitungsmechanismen nicht nur auf der Ebene der syntaktischen Verarbeitung wirken, sondern auch auf der Ebene referenzieller Verarbeitung.

### 7.1.2. Empirische Untersuchung zur Anaphernverarbeitung in LSK

Die untersuchten syntaktischen Strukturen beziehen sich direkt auf die Gegenstände der Simulationsstudien, die wir im vorausgegangenen Kapitel 6 vorgestellt haben (vgl. auch Weldle, Baumann et al., 2009). Die allgemeine Performanz der Modelle war auch für komplexe syntaktische Strukturen zufriedenstellend, strukturelle Constraints wurden korrekt erfüllt. Zugleich fanden wir aber für die Referenzzuweisungs-Komponente eine Sensitivität für den lokalen strukturellen Kontext, die zuvor für einfache SRNs für Aspekte der Verarbeitung syntaktischer Strukturen bzw. bei der Projektion der folgenden Wortklassen gefunden wurde (Konieczny, Ruh & Müller, 2009). In unseren Modellen haben wir den Einfluss von LSK auf die Bindungsdomänen von Personal- und Reflexivpronomen bzw. auf die Auswahl des Koreferenten überprüft. Die gefundenen Effekte hängen stark von den Bedingungen der gegebenen Stimuli ab: Während bei einer informationsreichen Grammatik mit satzinterner Interpunktion nur allgemeine Interferenz-Effekte auf die Anaphernaufflösung gefunden wurden, zeigte sich bei fehlender Interpunktion ein Effekt des syntaktischen Kontexts. Die Einbettung in verschiedene syntaktische Kontexte beeinflusst die Auswahl des Koreferenten der Anapher, wenn diese in einer Subsequenzen mit kohärenter Wortfolge und einem alternativen, lokalen Antezedens erscheint. Dabei wurden neben den korrekten Antezedenzen auch Referenten aktiviert, deren Auswahl einer strukturell verursachten Verletzung der global strukturellen Bindungsdomänen entspricht. Umgekehrt wurden die strukturellen Bindungsprinzipien innerhalb des lokalen Kontextes auch restriktiv wirksam: Im Fall des Personalpronomens wurde das vermeintliche Subjekt der lokalen Subsequenz nur reduziert aktiviert.

Gegenstand der im Folgenden präsentierten empirischen Untersuchungen sind – parallel zum Gegenstand der Modellierung – Personal- und Reflexivpronomen. Wie bereits näher erläutert wurde, wird in syntaktisch-strukturell orientierten Theorien (v.a. der Rektions-

und Bindungstheorie, Chomsky, 1980) eine exklusive komplementäre Bindungsdomäne angenommen (d.h. die möglichen Bezugsrahmen schließen sich gegenseitig aus), die aus der Anwendung der sich gegenseitig ausschließenden Prinzipien A und B resultieren.

**LSK als rein struktureller Einflussfaktor** Bezogen auf diese postulierten Prinzipien haben wir im vorausgegangenen Kapitel 5 verschiedene Studien zur Online-Verarbeitung von Reflexiv- und Personalpronomen besprochen, in denen empirische Befunde zu Verletzungen der Annahme einer strikten Komplementarität gezeigt wurden (vgl. u.a. Runner et al., 2003, 2006; Kaiser et al., 2009; MacWhinney, 2008), die systematisch unter bestimmten Voraussetzungen des syntaktischen oder pragmatischen Kontext auftreten. Besondere Aufmerksamkeit gilt in diesen Studien den oben erläuterten PNPs (Picture-Noun-Phrases, vgl. Kap. 5.3.2), für die bereits früh gezeigt werden konnte, dass die Bindung von Anaphern in diesem syntaktischen Kontext stärker durch semantisch-pragmatische Aspekte geleitet werden.

Die Satztypen, die wir unseren Studien zugrunde legen, sind explizit keine PNPs. Es werden ausschließlich Sätze verwendet, in denen im Fall der Personalpronomen eine transitive Handlung beschrieben wird – in denen ein Agens auf die durch das Personalpronomen referenzierte Person einwirkt. Im Fall der Reflexivpronomen wird eine reflexive Handlung des Agens beschrieben, bei der das Agens dementsprechend auf sich selbst einwirkt. Im Unterschied zu den vielbeforschten PNPs handelt es sich bei den von uns verwendeten Sätzen mit eingebetteten LSK also um ein explizit strukturelles Phänomen. Mit einem Befund der Verletzungen der Komplementaritätsannahme für diese Sätze weiten wir den Phänomenbereich aus, innerhalb dessen globale strukturelle Constraints keine absolute Gültigkeit besitzen.

Im Unterschied zu den Befunden von Runner et al. (2003, 2006) kann diese Verletzung allerdings nicht im Sinne von Logophoren erklärt und in den Bereich semantisch-pragmatischer Faktoren abgeschoben werden. Da unser Faktor struktureller Natur ist, müssen wir den Effekt anders interpretieren: Die Verletzung betrifft weniger die Annahme struktureller Constraints und des Komplementaritätsprinzips an sich, sondern stellt vielmehr die Annahme in Frage, dass strukturelle Constraints strikt auf der globalen Satzebene wirken. Dem entgegen wäre anzunehmen, dass diese in der Online-Verarbeitung unmittelbar zum Einsatz kommen, und – zumindest temporär – eine lokal korrekte statt einer global korrekten Bindung herstellen. Mögliche alternative Erklärungen werden zu diskutieren sein.

Die Feststellung, dass es sich bei LSK um einen strukturellen Faktor handelt, hat eine weitere Konsequenz: Den Annahmen des form-specific multiple-constraints approach (Brown-Schmidt et al., 2005; Kaiser et al., 2009) folgend, erwarten wir unterschiedliche Effekte für Personalpronomen gegenüber Reflexivpronomen. Wie Kaiser et al. (2009) zeigen, weisen verschiedene Anaphertypen eine spezifisch gewichtete Sensitivität für verschiedene Constraints (struktureller, syntaktischer, semantischer, pragmatischer Natur) auf (vgl. Kap. 5.3.2). Speziell Reflexivpronomen zeigen eine wesentlich höhere Sensitivität für strukturelle Constraints, während Prinzip B als Grundlage für die Filterung der möglichen Koreferenten von Personalpronomen als ausschließendes Prinzip nach dem Ausschluss des c-kommandierenden Agens keine weiteren strukturellen Hinweise auf das korrekte

Antezedens mehr zulässt. Für unsere Experimente heißt das, dass wir einen stärkeren Effekt der LSK für Reflexivpronomen erwarten, da er auf struktureller Ebene wirken und den Verarbeitungsmechanismen für reflexive Ausdrücke damit in stärkerem Umfang zugrunde liegen sollte als dies bei Personalpronomen der Fall ist.

Nicht zuletzt werden die Befunde natürlich in Beziehung zur Diskussion um den theoretischen Status lokaler syntaktischer Kohärenzen (Tabor et al., 2004; Konieczny, 2005; Konieczny, Müller, Hachmann et al., 2009; Konieczny, Müller, Baumann et al., 2009; Tabor et al., 1997; Demberg & Keller, 2009; Hale, 2011) zu setzen sein, da sie den Skopus dieses Untersuchungs-Paradigmas um die Domäne referenzieller Bindung erweitern.

**Empirischer Zugang** In der folgenden Experimentalserie verwenden wir verschiedene empirische Zugänge bzw. psycholinguistische Methoden zur Untersuchung unseres Gegenstands. Zum Einsatz kommen Lesezeitmessungen im Self-Paced Reading Paradigma mit anschließend zu beantwortenden geschlossenen Fragen, Blickbewegungs-Aufzeichnungen im Visual-World Paradigma und Ratings von auditiven Stimuli hinsichtlich ihrer Verständlichkeit und ihrer prosodischen Qualität. Dabei werden zwei Fragen zu klären sein: (a) Wie können wir mit diesen Methoden online die aktuelle Auswahl des Referenten auf der Position der anaphorischen Referenz messen, bzw. wie sicher können wir aus den Daten auf den gewählten Koreferenten schließen? (b) Welche Effekte der LSK erwarten wir in den verschiedenen Experimentalbedingungen? Die allgemein formulierten Annahmen werden in den spezifischen Hypothesen an das jeweilige Paradigma und die darin aufgezeichneten Performanz- und Verhaltensmaße angepasst werden.

## 7.2. AnaLoCo 1: Lesezeitmessung

Im ersten Experiment geht es zunächst darum herauszufinden, ob wir in der gegebenen Satzstruktur einen Interferenz-Effekt lokaler Kohärenz finden. Hierzu verwenden wir eine Lesezeit-Studie, wie sie beispielsweise von Tabor et al. (2004) durchgeführt wurde. Als Experimentalparadigma verwenden wir eine Self-Paced-Reading-Studie mit anschließend zu beantwortenden Fragen – Ja-Nein-Fragen (einschließlich Aufzeichnung der Beantwortungslatenz), die sich indirekt auf die Interpretation der anaphorischen Referenz bezieht. Ein Einfluss von LSK auf die Lesezeiten auf den kritischen Positionen unseres Materials gibt einen ersten Hinweis darauf, dass sich der lokale Kontext auf die betreffenden Positionen auswirkt und die Bindung von Anaphern an ihr Antezedens potenziell von solchen lokalen Verarbeitungsprozessen beeinflusst werden kann. Finden sich darüber hinaus Effekte auf die Latenz der Beantwortung der Frage sowie auf die Korrektheit der Antwort, weist dies auf einen tiefer greifenden Effekt des lokalen Kontexts auf die globale Interpretation hin. Über den zweiten Faktor vergleichen wir die Verarbeitung von Personal- und Reflexivpronomen. Wir erwarten eine Wechselwirkung zwischen den Faktoren Anapherentyp und LSK mit einem stärkeren Effekt der LSK-Bedingung für Reflexivpronomen. Eine solche Wechselwirkung zwischen den Personal- und Reflexivpronomen in ihrer Sensitivität für lokale Kohärenzen könnten im Sinne des Ansatzes formenspezifischer multipler Constraints interpretiert werden.

### 7.2.1. Design

Im Satzmaterial wurden zwei Faktoren mit jeweils zwei Bedingungen variiert – wir verwenden also ein  $2 \times 2$ -Design.<sup>2</sup> Wir unterscheiden die Faktoren Satzstruktur und Anapherotyp: Der Faktor Anapherotyp besteht aus den Ausprägungen Reflexivpronomen (*sich*) und Personalpronomen (*ihn*, immer im Genus Maskulinum und im Kasus Akkusativ). Der Faktor Satzstruktur erscheint in den Ausprägungen Struktur mit LSK und Kontrollstruktur mit Unterbindung der LSK. In der Kontrollstruktur wird eine potentielle LSK durch eine andere grammatische Struktur lokal inkohärent und damit uninterpretierbar gemacht. Die Korrektheit der Antwort wird als Pseudo-Faktor mit aufgenommen. Für beide Hauptfaktoren wird ein Einfluss sowohl auf die Lesezeiten als auch auf die Beantwortung und die Beantwortungslatenz der Fragen erwartet. Die Kodierung der Bedingungen kann Tabelle 7.1 entnommen werden, die Bedingungen werden in der folgenden Liste zusammengefasst.

Design		Faktor 1: Anapherotyp	
		sich	ihn
Faktor 2: LSK	mit LSK	1	2
	LSK durchbrochen	3	4

Tabelle 7.1.: Tabelle zum Design für Experiment 1.

**Bedingung 1** Reflexivpronomen (*sich*) mit lokal interferierender Lesart (pro LSK).

**Bedingung 2** Personalpronomen (*ihn*) mit lokal interferierender Lesart (pro LSK).

**Bedingung 3** Reflexivpronomen (*sich*), unterdrückte lokale Lesart (contra LSK).

**Bedingung 4** Personalpronomen (*ihn*), unterdrückte lokale Lesart (contra LSK).

### 7.2.2. Stimulus-Material

Das in der Studie verwendete Sprachmaterial hat folgende Struktur: Ein vorangestellter temporaler Nebensatz, der einen zentral eingebetteten Objekt-Relativsatz beinhaltet (d.h. einen Relativsatz, dessen Relativpronomen das Objekt des Relativsatzes ist – Struktur  $O_{\text{pro}}SV_{\text{fin}}$ ); dem Nebensatz folgt ein Hauptsatz.<sup>3</sup> Im Nebensatz schließt sich direkt an den Relativsatz eine anaphorische Referenz an: entweder ein Reflexivpronomen (Bedingungen 1 und 3) oder ein Personalpronomen (Bedingungen 2 und 4). Zwischen Anapher und Matrixverb des Nebensatzes folgt noch eine Präpositionalphrase. Um das Design transparent zu halten und die Bedingungen kontrollieren zu können, verwenden wir ausschließlich

<sup>2</sup>Wenn wir die Beantwortung der Fragen im Design berücksichtigen, haben wir genau genommen ein  $2 \times 2 \times 2$ -Design mit insgesamt 8 Bedingungen. Die Fragen beschreiben einen korrekten oder inkorrekten Sachverhalt und müssen mit *ja* oder *Nein* beantwortet werden. Eine entsprechende Kodierung haben wir für die Rotierung der Listen übernommen. Allerdings haben wir diese Kodierung in der Auswertung wiederum aufgehoben und die entsprechenden Fälle zusammengelegt, indem wir für die Fragen nur einen Faktor konstatieren, der die beiden Zustände *Frage korrekt beantwortet* oder *Frage falsch beantwortet* annehmen kann. Dies erleichtert die Auswertung bedeutend und macht die Ergebnisse transparenter.

<sup>3</sup>Der Hauptsatz ist für die kritischen Positionen der Studie irrelevant, spielt aber insofern eine Rolle, als auf Diskursebene weitere mögliche Referenten des Satzes eröffnet werden können.

Referenten im Genus Maskulinum, Numerus Singular. Satzobjekte und damit auch das Personalpronomen stehen immer im Kasus Akkusativ.

Diese Satzstruktur beinhaltet eine LSK, die über die Grenze des Relativsatzes hinausgeht, und lokal isoliert einem Hauptsatz entspricht: Die Subsequenz zieht sich vom Subjekt des Relativsatzes über das finite Verb des Relativsatzes und die Relativsatzgrenze bis zum Pronomen im Matrixsatz (sowohl global als auch lokal das Objekt des Satzes) und schließt das Ende der folgenden Präpositionalphrase mit ein (SV<sub>fin</sub>O<sub>pro</sub>PP). Im Beispiel entspricht dies der Subsequenz *der Sohn kämmt* (,) *sich im Wohnzimmer* bzw. *der Sohn kämmt* (,) *ihn im Wohnzimmer*. Die Einbettung in einen Nebensatz erzielt dabei durch die Verbletzstellung die Verschiebung des Matrixsatz-Verbs an das Ende des Satzes.

Im temporalen Nebensatz und im Relativsatz verwenden wir Verben, die zugleich reflexiv und transitiv verwendet werden können.<sup>4</sup> Dadurch erreichen wir, dass sich sowohl mit einem Reflexivpronomen als auch mit einem Personalpronomen ein grammatikalisch wie semantisch korrekter Satz ergibt – und zwar sowohl in der lokalen als auch der globalen Lesart.

Die LSK ist in den Bedingungen 1 und 2 gegeben. In den Kontrollbedingungen 3 und 4 wird die Kohärenz durch ein in den Objekt-Relativsatz vor das finite Verb eingefügtes Adverb durchbrochen. Dadurch besitzt die lokale Sequenz keine gültige Hauptsatzstruktur mehr und kann nicht als LSK gelesen werden.<sup>5</sup> Wir haben also gemäß unseres Designs vier Bedingungen: Die Zielbedingungen Reflexivpronomen in der LSK (Bsp. 26a) und Personalpronomen in der LSK (Bsp. 26b) sowie die Kontrollversionen Reflexivpronomen in durchbrochener LSK durch Adverbeinschub (Bsp. 26c) und Personalpronomen in durchbrochener LSK durch Adverbeinschub (Bsp. 26d).

- (26) a. Als der Vater<sub>i</sub>, den der Sohn<sub>j</sub> kämmt, sich<sub>i/\*j</sub> im Wohnzimmer anzieht, ist die halbe Stadt bereits auf dem Weg zur Aufführung.  
Frage: Zog sich der Vater an? (Antwort: Ja)  
Frage: Kämmte sich der Sohn? (Antwort: Nein)
- b. Als der Vater<sub>i</sub>, den der Sohn<sub>j</sub> kämmt, ihn<sub>\*i/j</sub>... im Wohnzimmer anzieht, ist die halbe Stadt bereits auf dem Weg zur Aufführung.  
Frage: Kämmte der Sohn den Vater? (Antwort: Ja)  
Frage: Zog der Sohn den Vater an? (Antwort: Nein)
- c. Als der Vater<sub>i</sub>, den der Sohn<sub>j</sub> gründlich kämmt, sich<sub>i/\*j</sub> im Wohnzimmer anzieht, ist die halbe Stadt bereits auf dem Weg zur Aufführung.  
Frage: Zog sich der Vater an? (Antwort: Ja)  
Frage: Kämmte sich der Sohn? (Antwort: Nein)
- d. Als der Vater<sub>i</sub>, den der Sohn<sub>j</sub> gründlich kämmt, ihn<sub>\*i/j</sub>... im Wohnzimmer anzieht, ist die halbe Stadt bereits auf dem Weg zur Aufführung.

<sup>4</sup>Da es im Deutschen nur sehr wenige Verben gibt, die dieses Kriterium erfüllen, haben wir darauf verzichtet, die Frequenz dieser Verben hinsichtlich ihrer reflexiven vs. transitiven Verwendung in die Auswahl mit einzubeziehen. Die Auswahl der verwendeten Verben erfolgte aus der subjektiven Einschätzung der Autoren.

<sup>5</sup>Dadurch wird wiederum eine neue lokal kohärente Struktur eröffnet: Ein Hauptsatz mit topikalisiertem Adverb (AV<sub>fin</sub>O<sub>pro</sub>PP...), entsprechend *gerade kämmt* (,) *sich im Wohnzimmer*... Diese Struktur ist im Deutschen allerdings sehr selten.

Frage: Kämmte der Sohn den Vater? (Antwort: Ja)

Frage: Zog der Sohn den Vater an? (Antwort: Nein)

Wie bereits oben angedeutet, wurde für die Rotation eine weitere Pseudo-Bedingung aufgenommen, die für die Auswertung wieder zusammengefasst wurde. In dieser Bedingung wird die an den Satz anschließende Frage spezifiziert. Es handelt sich um geschlossene Fragen (*Ja* vs. *Nein*), die sich direkt auf die Handlungen des Matrixsatzes (d.h., des temporalen Nebensatzes) oder des Relativsatzes beziehen. Dabei wird gefragt, ob Person A eine spezielle Handlung (Verb des Matrixsatzes oder des Relativsatzes) an sich selbst (in den reflexiven Sätzen) oder an Person B (in den transitiven Sätzen) ausführt. In der Frage wird eine Proposition abgefragt, die ein Argument enthält, auf das anaphorisch verwiesen wurde. Über diese Frage wird damit indirekt geprüft, welches Antezedens für die Anapher angenommen wurde, da sich die abgefragte Handlung auf den Koreferenten der Anapher bezieht. Durch die indirekte Abfrage der Wahl der Anaphernbindung wird auch vermieden, dass der Aspekt der Anaphernauflösung (als Gegenstand der Untersuchung) durch die Frage hervorgehoben wird. Für die Analyse wurden die Bedingungen dieses Faktors zusammengefasst, indem die konkreten Antworten mit der relativen Korrektheit der Antworten ersetzt wurden: Es wurde überprüft, ob die jeweilige Frage korrekt beantwortet wurde oder nicht.<sup>6</sup>

Es wurden in den vier Bedingungen je 24 Sätze erstellt, insgesamt also 96 Sätze. Dazu wurden korrespondierende Fragen in weiteren zwei Bedingungen erstellt, wodurch sich insgesamt 192 Satz-Frage-Paare ergeben.<sup>7</sup> Die Bedingungen für die 24 Experimental-Items wurden anhand eines Lateinischen Quadrats vollständig durchrotiert. Das Experiment wurde parallel mit zwei weiteren Self-Paced-Reading-Experimenten durchgeführt, deren Sätze eine vergleichbare Komplexität aufweisen, und deren Items als Füllmaterial für das Experiment dienten. In diesen Experimenten wurden jeweils 24 Experimental-Items (ebenfalls in 4 Bedingungen variiert) verwendet, insgesamt also 48 Sätze in 4 Bedingungen. Damit stehen Experimental-Items zu Füll-Items in unserem Experiment in einem Verhältnis von 1:2 (33% Experimental-Items).

### 7.2.3. Methodik und Hypothesen

In der ersten Studie wurde das Self-Paced Reading Moving Window-Paradigma (nicht-kumulativ) verwendet (vgl. Just et al., 1982; Tabor et al., 2004). Sätze wurden Wort für Wort per Tastendruck eingeblendet. Die einzelnen Wörter wurden durch unterstrichene *Xe* maskiert, beim Einblenden des jeweils nächsten Wortes wurde das vorangegangene Wort wieder remaskiert. Die Ansteuerung der Einblendung des jeweils nächsten Wortes erfolgte per Tastendruck (Leertaste der Tastatur). Wir haben die Methode um eine geschlossene Frage erweitert, die jeweils im Anschluss an den Satz durch Drücken der Taste *j* (für Ja) oder *n* (für Nein) zu beantworten war. Aufgezeichnet wurden die Latenz zwischen den

---

<sup>6</sup>Die vorausgegangene Unterscheidung der Antwortmöglichkeit wurde ausschließlich dazu benötigt, um eine vollständige Rotation der Bedingungen mit einer gleichmäßigen Verteilung der Antwortbedingungen zu erreichen.

<sup>7</sup>Die im Vorwort angekündigte elektronische Version des Anhangs im PDF-Format wird auf Anfrage (E-Mail an [helmut.weldle@googlemail.com](mailto:helmut.weldle@googlemail.com)) zugesandt. In diesem Anhang findet sich das vollständige Stimulus-Material des Experiments AnaLoCo 1.

Tastendrücken, die Antwort auf die geschlossene Frage sowie die Latenz nach Satzende bis zur Beantwortung der Frage.

Zur Darbietung der Stimuli und zur Aufzeichnung der Lesezeit-Daten haben wir das Package Linger (Rohde, 2001) verwendet.<sup>8</sup> Linger ist eine Sammlung von Methoden für computergestützte psycholinguistische Experimente, die in tcl/tk implementiert wurde und dadurch gut zu handhaben ist. Ausgeführt wurde das Experiment auf einem Standard-PC.<sup>9</sup>

### 7.2.3.1. Ablauf und Teilnehmer

Am Experiment nahmen 27 Versuchsteilnehmer (VTn) teil, vorwiegend Studierende der Universität Freiburg. Die Teilnehmer erhielten als Vergütung im Anschluss an das Experiment Versuchsteilnehmerstunden-Scheine<sup>10</sup> oder Geld (5,- Euro). Der Versuch dauerte ca. 30 Minuten.

Die VTn bekamen zu Beginn eine schriftliche Instruktion und konnten anschließend mit dem Versuchsleiter noch offene Fragen klären. Im Experiment wurden auf einem Computerbildschirm Sätze Wort-für-Wort dargeboten. Dabei wurde die Geschwindigkeit der Darbietung per Tastendruck durch den VTn selbst bestimmt. Hauptaufgabe der VTn war es, den jeweiligen Satz vollständig zu lesen. Als Sekundäraufgabe musste die anschließend erscheinende Frage (entsprechend der Aufforderung in der Instruktion) möglichst zügig beantwortet werden. Für die Beantwortung der Frage wurde keine Rückmeldung (korrekt oder falsch beantwortet) gegeben.<sup>11</sup> Nach einem Testlauf mit fünf Sätzen, nach denen nochmals Fragen an den Versuchsleiter gestellt werden konnten, folgte das Experiment mit den Experimental-Items.

### 7.2.3.2. Hypothesen

Unserem Experiment liegen folgende Hypothesen zugrunde: Sind Reflexiv- oder Personalpronomen in eine LSK eingebettet, in der alternative Anbindungsmöglichkeiten bestehen, sollte das lokale Agens als Antezedens in Betracht gezogen werden. Dabei muss die

---

<sup>8</sup>Die Verwendung von Linger ist zugegebenermaßen nicht ganz unproblematisch, da die ihm zugrundeliegende Programmiersprache *tcl/tk* eine interpretierte Sprache ist (Performanzverlust durch Übersetzung in Maschinensprache zur Laufzeit des Programms). Aufgrund des Zwischenspeichers der Tastatur oder der Tastaturanschlüsse tritt zusätzliches Rauschen in den Daten auf, was im schlimmsten Fall dazu führt, dass in den Daten vorhandene Effekte nicht gefunden werden. Eine ebenfalls kostenfreie Alternative wäre das Package DMDX/TimeDX, das eine wesentlich bessere zeitliche Auflösung hat, dafür aber nicht so schnell und einfach einsetzbar ist. Zudem wurde das Package Linger mittlerweile von mehreren Forschungsgruppen erfolgreich eingesetzt (vgl. z.B. Fedorenko, Gibson & Rohde, 2006; R. Levy, Fedorenko, Breen & Gibson, submitted).

<sup>9</sup>Stationärer PC, Intel Pentium Dual Core, 2.7 GHz, 2 GB RAM) mit USB-Tastatur und 22-Zoll TFT-Bildschirm.

<sup>10</sup>Im Rahmen des Studiums an der Universität Freiburg muss in den Fächern Kognitionswissenschaft und Psychologie für die Zulassung zur Zwischenprüfung eine bestimmte Anzahl an Versuchsteilnehmerstunden abgeleistet werden.

<sup>11</sup>Durch die Auslassung der Rückmeldung sollte vermieden werden, dass die VTn bei falscher Beantwortung über den Grund des Fehlers nachdenken und sich dadurch verfälschende Effekte im jeweiligen Folge-Trial ergeben.

Alternative im isolierten lokalen Kontext strukturell (strukturelle Bindungsprinzipien werden erfüllt), grammatisch (Kongruenz zwischen Anapher und Antezedens in Genus und Numerus) und semantisch (Plausibilität des Antezedens, z.B. mögliches belebtes Agens einer entsprechenden reflexiven Handlung) kompatibel sein. Wir nehmen also an, dass die global korrekte Zuweisung einer Anapher zu einem Antezedens temporär innerhalb einer LSK überschrieben werden kann: Ein lokales Antezedens, das global unzugänglich wäre, konkurriert mit dem global korrekten Antezedens.

Bei den Lesezeiten erwarten wir verschiedene Effekte und Interaktionen erleichterter und erschwelter Verarbeitung, die für den Faktor Satztyp während und direkt nach der LSK durch die lokale parallele Konstruktion alternativer syntaktischer Strukturen und für den Faktor Anapherentyp mit stärkerer Struktursensitivität des Reflexivpronomens zusammenhängen. Die Effekte sollten sich auf den im Folgenden benannten Positionen mit folgenden Ausprägungen finden:

1. Gesamter Verlauf der LSK, über den die LSK isoliert als Hauptsatz gelesen werden kann. Dabei können folgende Positionen hervorgehoben werden: Einstieg in die LSK am Verb (übereinstimmende Kongruenz zum lokalen Subjekt), anaphorische Referenz, anschließende Präpositionalphrase.
  - Einstieg in die LSK am finiten Verb des Relativsatzes: Aufgrund der einfacher zu verarbeitenden Hauptsatzstruktur erwarten wir, dass sich bereits am Verb ein entsprechender Effekt reduzierter Lesezeiten zeigt, da in der durchbrochenen Satz-Bedingung durch das Adverb die einfachere Hauptsatzstruktur bereits ausgeschlossen wird.
  - Anaphorische Referenz und Präpositionalphrase: Sollte die lokale Lesart als Hauptsatz noch bestehen, so sollte sich hier weiterhin eine geringere Lesezeit für die LSK-Bedingung zeigen.
2. Die Position der anaphorischen Referenz, auf der innerhalb der LSK lokale Constraints mit globalen Constraints konkurrieren.
  - Hier ist eine konkrete Vorhersage schwierig, da Effekte der Vereinfachung innerhalb des einfacher zu verarbeitenden Hauptsatzes (Vereinfachungen durch einen schnelleren Zugriff auf das präsentere lokale Antezedens) den umgekehrten Effekten der Erschwerung durch konkurrierende Constraints der Anapherauflösung entgegenstehen.
  - Vor allem erwarten wir eine Interaktion: Die Vereinfachung der LSK-Bedingung mit dementsprechend geringeren Lesezeiten sollte sich stärker auf das Reflexivpronomen auswirken, das stärker durch strukturelle Faktoren determiniert ist und deshalb eine höhere Sensitivität für die LSK aufweisen sollte.<sup>12</sup>
3. Auflösung der korrekten globalen Struktur nach dem Ende der LSK auf den ersten Wörtern nach der LSK.

---

<sup>12</sup>Zugegebenermaßen könnte sich hier aber auch ein genau gegenteiliger Effekt finden, der durch eine stärkere Konkurrenz der Antezedenzen für das Reflexivpronomen in zwei gegensätzlichen strukturell bedingten Alternativen zustande kommt. Unabhängig davon, welche Richtung der Effekt einschlägt, erwarten wir dennoch eine Interaktion.

- Dies betrifft die Position des finiten Verbs des Matrixsatzes, auf der die lokale Interpretation als falsch erkannt werden kann, sowie eventuell frühe Positionen des weiteren Satzes. Hier erwarten wir eine deutlich höhere Lesezeit für die Version mit LSK. Denkbar ist dabei auch eine Verzögerung des Effekts und ein entsprechendes Spillover auf die nächste Position.
4. Für die Korrektheit der Antworten sowie die Latenzen bei der Beantwortung (Offline-Maße) erwarten wir des Weiteren folgende Effekte für den Faktor Satztyp (mit vs. ohne LSK):
- Einerseits gehen wir davon aus, dass sich die LSK nur temporär auf die Interpretation auswirkt und anschließend der korrekte globale Zusammenhang wiederhergestellt wird. Dennoch ist es denkbar, dass die lokale Interferenz die Interpretation so gestört hat, dass die anschließende Beantwortung der Frage im Fall der Bedingung mit LSK durch die konkurrierenden Constraints der Bindung beeinträchtigt wird. Folglich erwarten wir in diesem Fall einen geringeren Anteil korrekt beantworteter Fragen für Sätze mit LSK.
  - Dementsprechend erwarten wir auch einen Unterschied in den Beantwortungslatenzen: Die Beeinträchtigung des Aufbaus der korrekten Anbindung könnte – unabhängig von der Korrektheit der Beantwortung – zu erhöhtem Aufwand beim Abruf der korrekten Proposition und somit zu höheren Latenzen in den Bedingungen mit LSK führen.

#### 7.2.4. Ergebnisse

Die deskriptive Statistik und die Inferenzstatistik wurden wiederum mit R durchgeführt (R Development Core Team, 2010). Nach einer kurzen Erläuterung der Aufbereitung der Daten besprechen wir zunächst die Ergebnisse anhand der Residuen der Lesezeiten für den gesamten Satz. Anschließend besprechen wir die Ergebnisse für einzelne Interest Areas (im Folgenden: IAs).

Für die Inferenzstatistik wurden gemischte lineare Modelle (linear mixed effects models, vgl. D. M. Bates, 2005; Baayen, Davidson & Bates, 2008; Baayen, 2008) verwendet.<sup>13</sup> Im Unterschied zu einfachen linearen Regressionsmodellen ermöglichen gemischte lineare Modelle den gleichzeitigen Einbezug von festen Faktoren (fixed effects) und Zufallsfaktoren (random effects). Sie stellen eine echte Alternative zu varianzanalytischen Verfahren und der Verwendung von getrennten F1- und F2-Analysen dar, die in der Psychologie und den Sozialwissenschaften klassischerweise eingesetzt werden. Mittlerweile zeichnet sich ein Trend ab, diese klassischen Verfahren zugunsten gemischter Modelle zu vermeiden. Diese bringen einige Vorteile mit sich (vgl. Baayen et al., 2008; Brysbaert, 2007):

- Die Vermeidung getrennter Analysen über Zufallsfaktoren verhindert den Verlust relevanter Daten für die Modellschätzung durch unterschiedliche Aggregation. Dies überwindet ein grundlegendes Problem der ANOVA, die eine getrennte Berech-

<sup>13</sup>Wir haben dafür die Funktionen *lmer* zur Berechnung der Modelle sowie die Funktionen *avplmer.fnc/pvals.fnc* zur Schätzung der Haupteffekte und Wechselwirkungen und zur Berechnung der p-Werte verwendet. Die Funktionen stammen aus dem R-Package *languageR* (Baayen, 2010).

nung für verschiedene Zufallsfaktoren erfordert (i.d.R. Aggregation über VTn und Items). Die getrennte Analyse nach F1 und F2 riskiert die Postulierung von Effekten, die eigentlich als Artefakte verworfen werden müssten, von Clark (1973) als „the language-as-fixed-effect fallacy“ bezeichnet.

- Die Möglichkeit, verschiedene Zufallsfaktoren gemeinsam in ein Modell zu integrieren, erlaubt es, beliebige relevante Prädiktoren in die Datenmodellierung einzubeziehen. So können beispielsweise Reaktionszeiten in Beziehung gesetzt werden zur Korrektheit der Reaktion (z.B. bei der Beantwortung von Fragen), zu anderen Reaktionszeiten (z.B. zur Lesezeit des vorausgegangenen Items, um Spillover-Effekte zu berücksichtigen), oder zu Übungs- oder Ermüdungseffekten, die sich im Laufe eines Experiments einstellen können.
- Gemischte Modelle erlauben die Erfassung komplexer Designs und haben eine größere Flexibilität bei der Berechnung von Zufallseffekt-Strukturen.
- Gemischte Modelle sind im Unterschied zu ANOVAs robust gegenüber Verletzungen von Verteilungsannahmen. Für die Verwendung einer ANOVA müssen Normalverteilung und Homoskedastizität der Daten und Vollständigkeit des Designs gegeben sein.<sup>14</sup> Gemischte Modelle sind robust gegenüber fehlenden Messwerten und nicht normalverteilten Daten.
- Gemischte Modelle weisen unter Umständen eine höhere Teststärke auf (vgl. auch T. F. Jaeger, 2008).

### 7.2.4.1. Aufbereitung der Daten

Zur Vorbereitung des Datensets und zur Bereinigung von Ausreißern wurden mehrere Schritte durchgeführt: Definition von Interest Areas (relevante Satzpositionen), Logarithmierung der Lesezeiten, Bereinigung nach Outliern, Berechnung der Residuen sowie die Feststellung der Rate korrekt beantworteter Fragen.

Zunächst haben wir IAs definiert und diesen relevante Einzelwörter und Wortgruppen zugewiesen. Von besonderem Interesse sind dabei die Positionen innerhalb der LSK und in deren unmittelbarem Anschluß: (a) finites Verb des Relativsatzes, (b) anaphorischer Ausdruck, (c) Präpositionalphrase nach der Anapher, (d) finites Verb des Matrixsatzes, (e) die Folgepositionen nach dem LSK-auflösenden Matrixverb, falls es zu Spillover-Effekten kommen sollte. In IAs, die mehrere Wörter zusammenfassen (z.B. die Präpositionalphrase), wurde der Mittelwert von den Lesezeiten der Einzelwörter verwendet. Um die Linksschiefe der Verteilung auszugleichen, die bei Lesezeiten üblicherweise zu finden ist, wurden die Lesezeiten für die IAs logarithmiert.

Für die Bereinigung der Lesezeit-Daten wurde ein gemischtes lineares Modell auf die Rohdaten gefittet. Hierfür haben wir die logarithmierten Lesezeiten auf den IAs verwendet. In das Modell gingen als feste Faktoren die Länge der Interest Area (Anzahl der Buchstaben) und die Experimentalbedingungen ein. Als Intercept ging die Identität des Satzes ein, als Random Slope wurde die Itemposition in Abhängigkeit der Teilnehmer verwendet, um

---

<sup>14</sup>D.h. alle Zellen des Designs (nach Rotation) müssen besetzt sein. Der Normalfall psychologischer Forschung hingegen sind Datensets mit fehlenden Daten.

individuelle Übungs- und Ermüdungseffekte zu erfassen.<sup>15</sup> Zur Feststellung von Outliern wurden Boxplots verwendet. Dabei wurden in Abhängigkeit der festen Faktoren alle Datenpunkte entfernt, die den Abstand von 1.5 Standardabweichungen vom arithmetischen Mittel überschreiten. Mit dieser Methode wurden insgesamt 190 Datenpunkte entfernt. Bei einer Anzahl von 7250 Datenpunkten im Ausgangs-Datensatz sind das 2.62% der Daten.

Für die statistische Analyse wurden aus dem resultierenden Datensatz für die logarithmierten Lesezeiten auf den IAs die Residuen berechnet. Als zusätzliche Informationsquelle wurden Spillover-Kovariaten ermittelt. Dafür wurden für jedes Wort die Residuen der logarithmierten zentrierten Lesezeiten der Wörter auf den zwei vorausgegangenen Satzpositionen mit in den Datensatz aufgenommen. Die Berücksichtigung der Spillover-Kovariaten dient dem Ausgleich des Effekts vorhergehender Lesezeiten auf die aktuelle Lesezeit und der Erfassung der auf das Folgeelement überlagerten Komplexität (vgl. T. F. Jaeger, 2008; T. F. Jaeger et al., submitted).<sup>16</sup>

Der Anteil korrekt beantworteter Fragen ging nicht als Auswahlkriterium in die Datenbereinigung mit ein. Es wurden auch keine Items in Abhängigkeit der korrekten Beantwortung entfernt. Über 70% der Fragen wurden korrekt beantwortet, dabei lag der Anteil korrekt beantworteter Daten bei keinem der VTn unter 50%. Die Beantwortung scheint mit einigen Schwierigkeiten behaftet gewesen zu sein und weist auf eine hohe Komplexität der zu verstehenden Sätze hin. Dies wird in der späteren Diskussion zu berücksichtigen sein.

#### 7.2.4.2. Residuen der logarithmierten Lesezeiten

Die folgenden Abbildungen 7.1, 7.2 und 7.3 zeigen Liniendiagramme für die Residuen der logarithmierten Lesezeiten.<sup>17</sup> Abbildung 7.1 erfasst die Daten in Abhängigkeit aller vier Experimentalbedingungen. Die folgenden Abbildungen zeigen die Daten getrennt nach Experimentalfaktor: Abbildung 7.2 erfasst den Unterschied zwischen den Bedingungen des Faktors Satztyp (LSK und Kontrolle), Abbildung 7.3 den Unterschied zwischen Reflexiv- und Personalpronomen (Faktor Anapherntyp). Wir besprechen zuerst die einfacher nachvollziehbaren Abbildungen getrennt nach Faktoren und gehen dann auf mögliche Interaktionen ein, die in der gemeinsamen Abbildung ersichtlich werden.

Diagramm 7.2 zeigt die Residuen für die logarithmierten Lesezeiten auf den IAs für den Faktor Satztyp mit den Bedingungen LSK und Kontrolle. Hinsichtlich der Kontrastierung der Bedingungen wird es nach der Satzposition auf dem Adverb (*gründlich*) interessant. Diese ist nur in der Kontrollbedingung gegeben. Die erste Position, auf der sich ein Unterschied zeigen könnte, ist also das finite Verb des Relativsatzes (*kämmt*). Die hinsichtlich der einfacheren lokalen Hauptsatzstruktur erwartete Beschleunigung der Lesezeit für die

<sup>15</sup>Gemischtes lineares Modell auf Rohdaten zur Identifizierung von Outliern:

Linear mixed model fit by REML

Formula:  $\log.time.ia \sim ia.len + Condition + (1 | item) + (1 + item.pos | subject)$

<sup>16</sup>Zum Teil kommt der Anteil der benötigten Verarbeitungszeit für ein Wort erst später zum Tragen (Spillover). Dieser Spillover-Effekt führt zu höheren Lesezeiten auf späteren Wörtern, und kann diesen späteren Wörtern folglich nicht zugerechnet werden. Durch Einführung der vorhergehenden Lesezeiten als Kovariate kann dieser Effekt ausgeglichen werden. Die Funktion zur Outlier-Identifizierung und -Bereinigung und zur Feststellung der Spillover-Kovariaten wurden von Sascha Wolfer zur Verfügung gestellt. Vielen Dank an dieser Stelle.

<sup>17</sup>Die Abbildungen wurden mit der Funktion *lineplot.CI* im R-Package *sciplot* (Morales, 2010) erstellt.

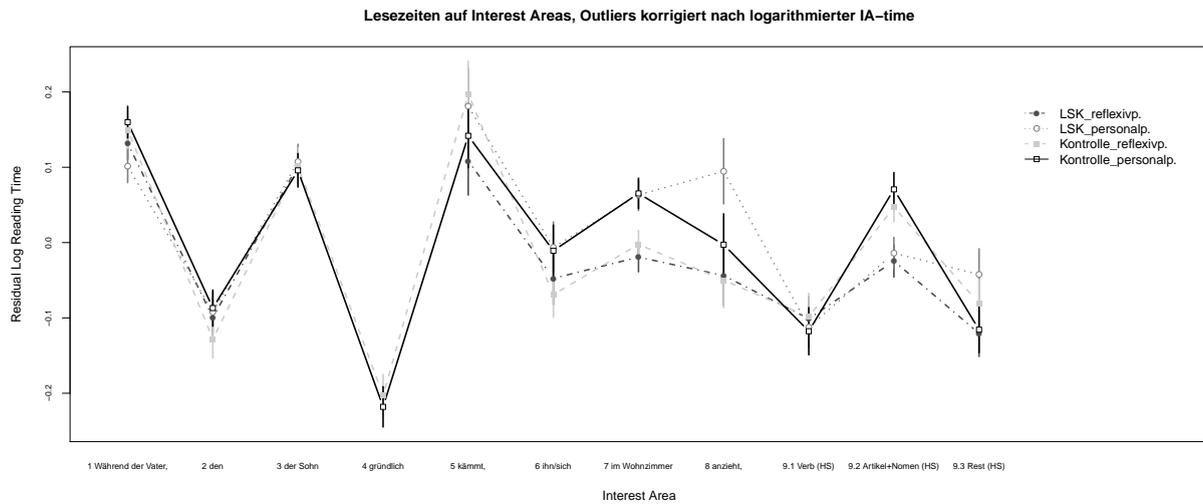


Abbildung 7.1.: AnaLoCo 1. Liniendiagramm der Lesezeitmessungen: Auf der x-Achse sind die IAs der Experimentalsätze abgetragen, auf der y-Achse die Residuen aus den logarithmierten Lesezeiten. Die Linien tragen die Lesezeiten auf den IAs für alle vier Bedingungen ab.

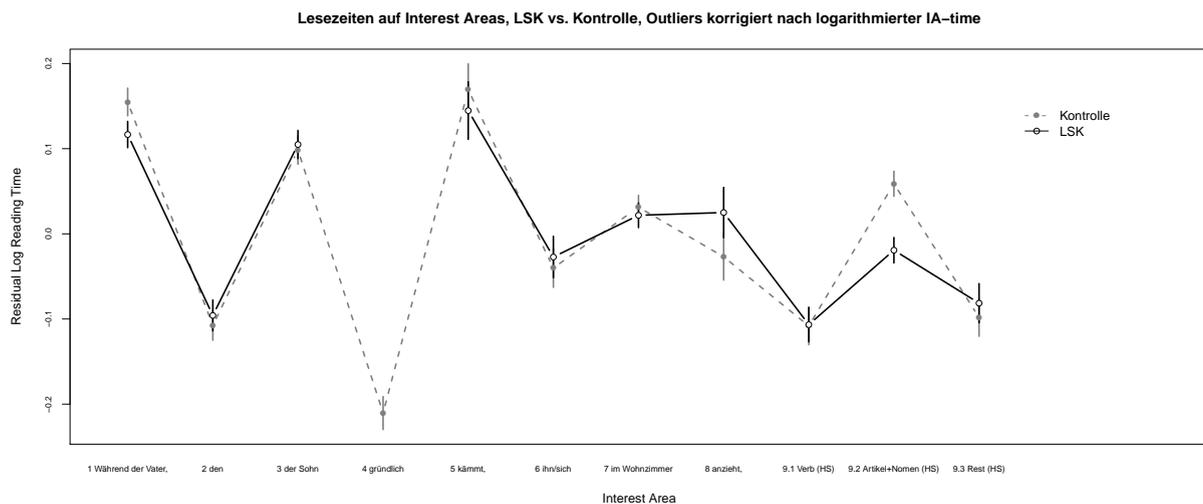


Abbildung 7.2.: AnaLoCo 1. Liniendiagramm der Lesezeitmessungen: Residuen aus den logarithmierten Lesezeiten auf den IAs für den Faktor Satztyp (LSK vs. Kontrolle).

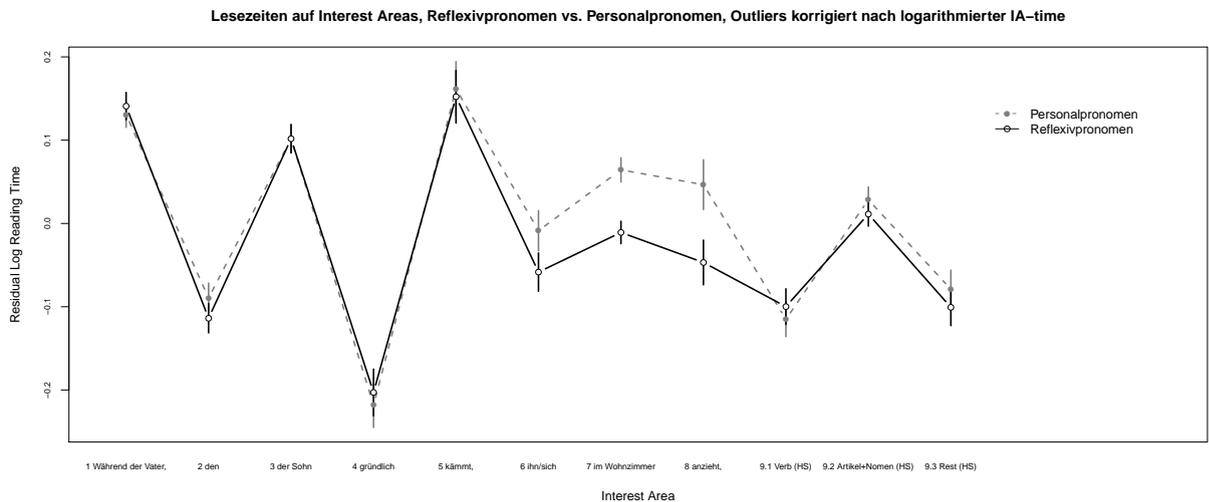


Abbildung 7.3.: AnaLoCo 1. Liniendiagramm der Lesezeitmessungen: Residuen aus den logarithmierten Lesezeiten auf den IAs für den Faktor Anapherntyp (Reflexivpronomen vs. Personalpronomen).

LSK-Bedingung zeigt sich hier nur als vage Tendenz. Der Effekt ist nicht signifikant. Das gleiche gilt auch für die folgenden Positionen auf dem anaphorischen Ausdruck (*sich/ihn*) und der Präpositionalphrase (*im Wohnzimmer*). Ein erster Unterschied deutet sich auf der Position des finiten Verbs des Matrixsatzes an (*anzieht*): An dieser Stelle sollte im Fall der tatsächlichen Berücksichtigung der LSK eine deutliche Verzögerung der Lesezeit im Vergleich zur Kontrollversion eintreten, da an dieser Stelle erkennbar wird, dass das neue Wort nicht sinnvoll in den lokalen syntaktischen Kontext eingebettet werden kann. Tatsächlich weist die Tendenz in die erwartete Richtung. Auch für die folgende Position wurde ein Unterschied im Sinne eines Spillover-Effekts gefunden. Dem Diagramm können wir aber entnehmen, dass sich dieser Effekt wesentlich stärker verzögert und erst auf der Position der beiden Wörter nach dem finiten Verb des beginnenden Hauptsatzes erscheint. Diese Position enthält keine einheitliche Klasse von Wörtern oder Konstituenten mehr, besteht aber zumeist aus einem Artikel und einem Nomen oder aus einem Artikel und einem Adjektiv/Adverb. Der Effekt weist in eine unerwartete Richtung – er deutet eine geringere Lesezeit im Fall der LSK-Bedingung an.

Diagramm 7.3 zeigt die Residuen für die logarithmierten Lesezeiten auf den IAs für den Faktor Anapherntyp mit den Bedingungen Reflexivpronomen und Personalpronomen. Für diesen Faktor können wir keine konkreten Erwartungen formulieren, da sich unsere Hypothesen auf die Interaktion zwischen den Faktoren Satztyp und Anapherntyp beziehen. Da die Auflösung von Reflexivpronomen wesentlich durch strukturelle Constraints bedingt und für die Zuweisung zu einem konkreten Antezedens nicht so sehr auf weitere Constraints angewiesen ist, können wir mit aller Vorsicht einen leichten Vorteil für die Version mit Reflexivpronomen erwarten. Die erste Position, auf der wir einen Unterschied erwarten können, ist die Position der anaphorischen Referenz (*sich/ihn*), denn erst ab hier unterscheiden sich die Bedingungen des Faktors Anapherntyp. Tatsächlich finden wir ab dieser

Position einen Unterschied in erwarteter Richtung, der sich über die Präpositionalphrase (*im Wohnzimmer*) bis zum finiten Verb des Matrixsatzes (*anzieht*) zieht.

Für Hinweise auf mögliche Interaktionen konsultieren wir Diagramm 7.1 mit den Residuen für die logarithmierten Lesezeiten über den IAs für alle vier Bedingungen.<sup>18</sup> Obwohl wir auf allen vier Positionen von der Anapher (*sich/ihn*) bis zur frühen Position des folgenden Hauptsatzes (Position 9.2, *Artikel + Nomen HS*) eine Interaktion erwartet haben, die auf eine stärkere Auswirkung der Bedingung LSK auf die Bedingung Reflexivpronomen hinweist, ergibt sich nur auf der Position des finiten Verbs des Matrixsatzes (*anzieht*) ein Hinweis auf eine Interaktion – allerdings nicht in der erwarteten Richtung. Des Weiteren ist auffallen, dass sich die Linien zwischen den Positionen des Relativsatzverbs und der Anapher durch die Bedingung Reflexivpronomen im Kontrollsatz kreuzen. Der Wechsel dieser Bedingung von der höchsten Latenz auf der ersten zur niedrigsten Latenz auf der zweiten Position gibt einen Hinweis darauf, dass es notwendig ist, eine Spillover-Kovariate bei den folgenden Berechnungen der Modelle für die inferenzstatistische Analyse miteinzubeziehen.

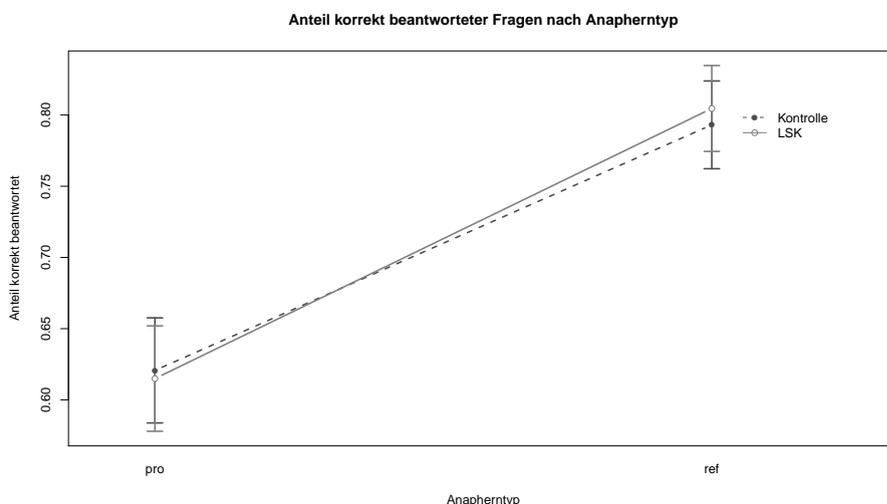


Abbildung 7.4.: AnaLoCo 1. Liniendiagramm des Anteils korrekt beantworteter Fragen: Auf der x-Achse sind die Anaphern (Reflexivpronomen vs. Personalpronomen) abgetragen, auf der y-Achse der Anteil korrekt beantworteter Fragen. Die Linien tragen die Bedingungen des Faktors Satztyps ab.

Die Abbildungen 7.4 und 7.5 zeigen Liniendiagramme für den Anteil korrekt beantworteter Fragen und die Latenzen bei der Beantwortung für die vier Bedingungen. Bei der Korrektheit zeigt sich kein Effekt für den Faktor Satztyp. Für die Latenzen zeichnet sich jedoch eine Tendenz für diesen Faktor ab. In beiden Fällen zeigt sich deutlich ein Effekt für den Faktor Anapherotyp.

<sup>18</sup>Für die Satztyp-Bedingung LSK werden die Mittelwerte durch Kreise dargestellt, die Bedingung Reflexivpronomen erhält in der LSK-Bedingung eine dunkelgraue gestrichelte Linie mit gefüllten Kreisen, die Bedingung Personalpronomen erhält eine hellgraue gepunktete Linie mit offenen Kreisen. Für die Satztyp-Bedingung Kontrolle werden die Mittelwerte mit Quadraten dargestellt, die Bedingung Reflexivpronomen wird hier durch eine hellgraue gestrichelte Linie mit gefüllten Quadraten repräsentiert, die Bedingung Personalpronomen durch die schwarze durchgezogene Linie mit offenen Quadraten.

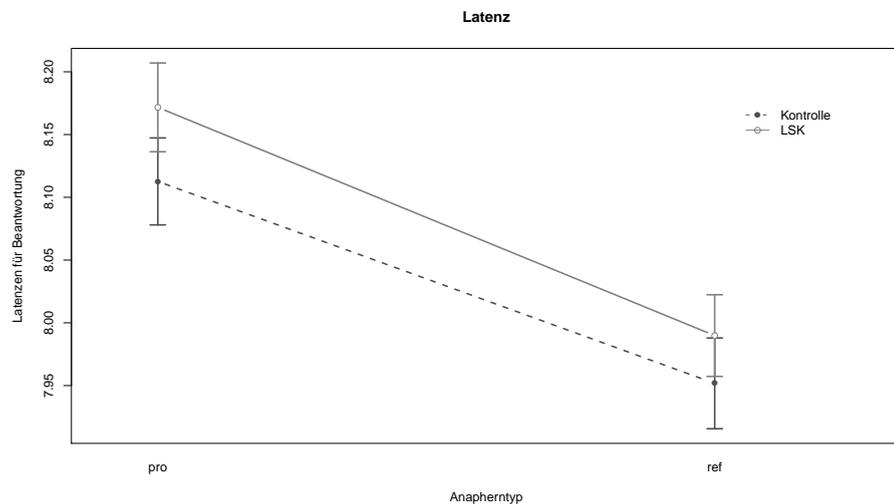


Abbildung 7.5.: AnaLoCo 1. Liniendiagramm der Reaktionszeiten zur Beantwortung der Fragen: Auf der x-Achse sind die Anaphern (Reflexivpronomen vs. Personalpronomen) abgetragen, auf der y-Achse die logarithmierten Beantwortungszeiten. Die Linien tragen die Bedingungen des Faktors Satztyps ab.

### 7.2.4.3. Gemischte lineare Modelle für IAs

Für folgende fünf Positionen wurden gemischte lineare Modelle für die Daten gefittet: (a) finites Relativsatz-Verb, (b) anaphorische Referenz, (c) Mittelwert über der Präpositionalphrase, drei Wörter, (d) finites Matrixsatz-Verb, (e) Position nach dem topikalisierten finiten Verb des folgenden Hauptsatzes, zwei Wörter. Des Weiteren wurde ein gemischtes lineares Modell für die Latenz bei der Beantwortung der Frage sowie ein gemischtes Log-it Modell für binomial verteilte Daten für die Korrektheit der beantworteten Fragen gefittet.

Beim Fitten der einzelnen Modelle wurden jeweils verschiedene feste Faktoren (fixed effects) und Zufallsfaktoren (random effects) aufgenommen. Als feste Faktoren wurden stets die Experimentalfaktoren Anapherentyp und Satztyp als Haupteffekte und Wechselwirkungen aufgenommen. Hinzu kamen jeweils verschiedene weitere feste Faktoren und Zufallsfaktoren (random effects). Das Modell mit dem besten Fit wurde ausgewählt.<sup>19</sup> Folgende zusätzliche feste Faktoren und Zufallsfaktoren wurden berücksichtigt:

- Als fester Faktor die Spillover-Kovariaten des ersten ( $w - 1$ ) und zweiten ( $w - 2$ ) vorausgehenden Wortes.
- Als fester Faktor der Logarithmus der Position des Experimental-Items im Verlauf des Experiments (vgl. z.B. T. F. Jaeger et al., submitted) zur Erfassung von Geschwindigkeitsveränderungen im Verlauf des Experiments.
- Als fester Faktor die Korrektheit der Antwort (binär kodiert: 0 für falsch beantwortete Fragen, 1 für richtig beantwortete Fragen).

<sup>19</sup>Der Vergleich der kontrastierenden Modelle erfolgte über paarweise durchgeführte ANOVAs. Das jeweils verwendete Modell wird als Funktionstherm dargestellt.

- Als Zufallsfaktor der Intercept für VTn: Lesezeiten sind global abhängig von der Person, die den Satz liest – interindividuelle Differenzen führen zu verschiedenen Lesezeiten und Latenzen). Durch die Aufnahme dieses Zufallsfaktors werden die interindividuellen Unterschiede im Modell berücksichtigt.
- Als Zufallsfaktor der Intercept für Items: Im Set der Experimental-Items kann es aufgrund minimaler semantischer Unterschiede bessere und schlechtere Items geben. Eine absolute Homogenität lässt sich hier nicht vollständig und endgültig gewährleisten. Die Aufnahme dieses Zufallsfaktors gleicht diese Unterschiede aus.
- Als Zufallsfaktor der Slope für die logarithmierte Item-Position in Abhängigkeit des VTn: Im Verlauf eines Experiments kann es bei längerer Dauer zu Ermüdungseffekten (größere Latenzen) und/oder zu Übungseffekten (geringere Latenzen) kommen. Die Ausprägung solcher Effekte während des Experimentverlaufs ist abhängig von der Person, die das Experiment bearbeitet (auch hier finden sich interindividuelle Unterschiede). Um solche Effekte im Modell zu berücksichtigen, wird ihr Slope mit in die Modellschätzung aufgenommen.
- Als Zufallsfaktor der Slope für die Korrektheit der Antwort in Abhängigkeit der VTn, um den Einfluss des individuellen Leseverständnisses auf die Lesegeschwindigkeit und die Antwortlatenz zu erfassen.

Die gefitteten Modelle werden in den folgenden Tabellen aufgeführt. Wir zeigen dabei ausschließlich für diejenigen Positionen Tabellen, auf denen Effekte gefunden wurden. Positionen mit nicht signifikanten Ergebnissen werden in Kurzform abgehandelt. Die Formel gibt die festen Faktoren und die Zufallsfaktoren wieder, die dem Modell zugrunde liegen und anhand derer der beste Fit erzielt wurde. Die folgenden Tabellen legen die Effekte der festen Faktoren und der Zufallsfaktoren dar. Die abschließende Tabelle führt die Ergebnisse im Format einer Varianzanalyse für die Haupteffekte auf. Hierfür wurde die Funktion *aovlmer.fnc* verwendet, die neben F-Werten auch eine Schätzung für p-Werte und Freiheitsgrade zurückgibt.<sup>20</sup>

**Relativsatz-Verb** Für das finite Verb des Relativsatzes wurde kein Effekt gefunden. Erwartet wurde ein Effekt für die Bedingung Satztyp (eine Beschleunigung in der LSK-Bedingung). Dies wurde durch das gefittete Modell<sup>21</sup> aber nicht bestätigt (Satztyp:  $t = -0.682$ ,  $F = 0.4989$ ,  $p > 0.48$ , Anapherntyp:  $t = 0.9$ ,  $F = 0.1213$ ,  $p > 0.72$ , Interaktion Satztyp:Anapherntyp:  $t = -1.652$ ,  $F = 2.7297$ ,  $p > 0.09$ ).

---

<sup>20</sup>Standardmäßig werden durch die Funktion *lmer* t-Werte ausgegeben, nicht aber Freiheitsgrade und p-Werte. Der Grund hierfür ist, dass nach aktuellem Forschungs- und Entwicklungsstand noch nicht geklärt ist, wie die Freiheitsgrade berechnet werden müssen (vgl. Baayen, 2008). Um p-Werte zu erhalten (sofern man das möchte, um eine Vergleichbarkeit mit traditionell ermittelten Signifikanzniveaus herzustellen), kann die Funktion *pvals.fnc* verwendet werden, die mit Monte-Carlo-Markov-Ketten-Simulationen (MCMC-Samples) eine Schätzung des p-Werts ausgeben. Die Funktion *aovlmer.fnc* verwendet ebenfalls MCMC-Samples zur Berechnung der p-Werte und gibt zudem Werte für Interaktionen aus. Dies ist bei der Funktion *pvals.fnc* nicht der Fall. Die Schätzungen sind eher anti-konservativ und vor allem hinsichtlich der ermittelten Freiheitsgrade mit Vorsicht zu genießen.

<sup>21</sup>Linear mixed model fit by REML:

Formula: `res.log.time.ia ~ loco * anaph + time.left + time.2left + log(item.pos) + (1 | item) + (1 + log(item.pos) | subject)`

**Personal-/Reflexivpronomen** Auch für die Position der anaphorischen Referenz wurden keine Effekte und auch keine Interaktion gefunden – weder für den Faktor Satztyp, noch für den Faktor Anapherntyp. Erwartet wurde eine Beschleunigung der Lesedauer für die Bedingung LSK sowie eine Interaktion mit dem Faktor Anapherntyp in Richtung einer stärkeren Effektauswirkung für Reflexivpronomen. Doch auch hier wurden die Erwartungen im Modell<sup>22</sup> nicht bestätigt (Satztyp:  $t = 0.775$ ,  $F = 0.1336$ ,  $p > 0.71$ , Anapherntyp:  $t = -1.072$ ,  $F = 2.6592$ ,  $p > 0.1$ , Interaktion Satztyp:Anapherntyp:  $t = -0.299$ ,  $F = 0.0893$ ,  $p > 0.76$ ).

**Präpositionalphrase** Für den Mittelwert der logarithmierten Residuen auf der Position der Präpositionalphrase (arithmetisches Mittel über drei Wörter) wurde schließlich ein höchst signifikanter Haupteffekt für den Faktor Anapherntyp gefunden ( $t = -2.352$ ,  $F = 17.432$ ,  $p < 0.0001$ ). Dieser bestätigt den Befund der geringeren Lesezeit für Reflexivpronomen aus Abbildung 7.3 (vgl. Tab. 7.2). Dies entspricht dem erwarteten Unterschied zwischen den Anapherntypen. Allerdings wurde damit weder die Erwartung eines Effekts der Bedingung Satztyp noch die Erwartung einer Interaktion von Satztyp und Anapherntyp im Sinne einer stärkeren Auswirkung der Verarbeitung von LSK auf Reflexivpronomen bestätigt (Satztyp:  $t = 0.16$ ,  $F = 0.2055$ ,  $p > 0.65$ , Interaktion Satztyp:Anapherntyp:  $t = -0.651$ ,  $F = 0.4238$ ,  $p > 0.51$ ).

**Matrixsatz-Verb** Auch für die Position des finiten Matrixsatz-Verbs wurde ein höchst signifikanter Haupteffekt des Faktors Anapherntyp gefunden (längere Lesedauer für Personalpronomen, vgl. Tab. 7.3). Hinzu kommt eine schwache Tendenz für den Faktor Satztyp – gemäß den Erwartungen mit längerer Lesedauer für die LSK-Bedingung (Anapherntyp:  $t = -1.217$ ,  $F = 10.9448$ ,  $p < 0.001$ , Satztyp:  $t = 1.99$ ,  $F = 3.0334$ ,  $p = 0.082$ ; bei getrennter Auswertung des Faktors Satztyp für die Bedingung Personalpronomen:  $t = 1.757$ ,  $F = 3.4813$ ,  $p = 0.062$ , vgl. auch Befunde nach Abb. 7.1). Allerdings konnte keine Interaktion zwischen den Experimentalfaktoren festgestellt werden ( $t = -1.062$ ,  $F = 1.1272$ ,  $p > 0.28$ ).

**Nachfolgender Hauptsatz** Erst auf der Position für die folgenden beiden Wörter nach dem Verb des nachfolgenden Hauptsatzes finden wir die erwarteten Effekte – Allerdings nicht in der Ausprägung, die vorhergesagt wurde (vgl. Befunde nach Abb. 7.2): Es zeigt sich ein höchst signifikanter Haupteffekt für den Faktor Satztyp ( $t = -3.141$ ,  $F = 21.1872$ ,  $p < 0.0001$ ). Dieser zeigt aber eine geringere Lesedauer für die Bedingung LSK an. Für den Faktor Anapherntyp zeigt sich kein Effekt und auch keine Interaktion (Anapherntyp:  $t = -0.640$ ,  $F = 0.8968$ ,  $p > 0.34$ , Interaktion Satztyp:Anapherntyp:  $t = 0.06$ ,  $F = 0.0036$ ,  $p > 0.9$ ).

<sup>22</sup>Linear mixed model fit by REML,

Formula: res.log.time.ia ~loco \* anaph + time.left + (1 | item) + (1 | subject)

Linear mixed model fit by REML

Formula: res.log.time.ia ~loco \* anaph + time.left + time.2left + log(item.pos) + (1 | item) + (1 + log(item.pos) | subject)

Random effects:

Groups	Name	Variance	Std.Dev.	Corr
subject	(Intercept)	0.0414819	0.203671	
	log(item.pos)	0.0022289	0.047211	-0.917
item	(Intercept)	0.0040559	0.063686	
Residual		0.0542225	0.232857	

Number of obs: 689, groups: subject, 29; item, 24

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	8.621e-02	5.685e-02	1.516
loco1	4.088e-03	2.558e-02	0.160
anaphref	-5.972e-02	2.540e-02	-2.352
time.left	4.440e-05	2.571e-05	1.727
time.2left	-2.100e-05	1.595e-05	-1.316
log(item.pos)	-7.454e-03	1.386e-02	-0.538
loco1:anaphref	-2.338e-02	3.590e-02	-0.651

Analysis of Variance Table:

	Df	Sum Sq	Mean Sq	F value	F	Df2	p
loco	1	0.01114	0.01114	0.2055	0.2055	682	0.65044
anaph	1	0.94520	0.94520	17.4320	17.4320	682	0.00003
time.left	1	0.24149	0.24149	4.4536	4.4536	682	0.03519
time.2left	1	0.08802	0.08802	1.6234	1.6234	682	0.20305
log(item.pos)	1	0.01393	0.01393	0.2569	0.2569	682	0.61245
loco:anaph	1	0.02298	0.02298	0.4238	0.4238	682	0.51524

Tabelle 7.2.: AnaLoCo 1. Gemischtes lineares Modell für die Residuen der logarithmierten Lesezeiten auf der Position der Präpositionalphrase (Mittelwert der Wortlesezeiten über die Präpositionalphrase).

Linear mixed model fit by REML

Formula:  $\text{res.log.time.ia} \sim \text{loco} * \text{anaph} + \text{time.left} + \text{time.2left} + \text{log(item.pos)} + (1 | \text{item}) + (1 + \text{log(item.pos)} | \text{subject})$

Random effects:

Groups	Name	Variance	Std.Dev.	Corr
subject	(Intercept)	0.019267	0.13881	
	log(item.pos)	0.006309	0.07943	-0.221
item	(Intercept)	0.015835	0.12584	
Residual		0.174542	0.41778	

Number of obs: 655, groups: subject, 29; item, 24

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	-7.265e-02	8.283e-02	-0.877
loco1	9.362e-02	4.705e-02	1.990
anaphref	-5.715e-02	4.698e-02	-1.217
time.left	2.069e-04	4.234e-05	4.886
time.2left	-7.870e-05	4.424e-05	-1.779
log(item.pos)	2.616e-02	2.467e-02	1.060
loco1:anaphref	-6.994e-02	6.588e-02	-1.062

Analysis of Variance Table:

	Df	Sum Sq	Mean Sq	F value	F	Df2	p
loco	1	0.5295	0.5295	3.0334	3.0334	648	0.082041
anaph	1	1.9103	1.9103	10.9448	10.9448	648	0.000991
time.left	1	3.7559	3.7559	21.5188	21.5188	648	0.000004
time.2left	1	0.6088	0.6088	3.4883	3.4883	648	0.062256
log(item.pos)	1	0.2237	0.2237	1.2815	1.2815	648	0.258043
loco:anaph	1	0.1967	0.1967	1.1272	1.1272	648	0.288779

Tabelle 7.3.: AnaLoCo 1. Gemischtes lineares Modell für die Residuen der logarithmierten Lesezeiten auf der Position des finiten Verbs des Matrixsatzes.

Linear mixed model fit by REML

Formula: res.log.time.ia ~ loco \* anaph + time.left + log(item.pos) + (1 | item) + (1 + log(item.pos) | subject)

Random effects:

<i>Groups</i>	<i>Name</i>	<i>Variance</i>	<i>Std.Dev.</i>	<i>Corr</i>
subject	(Intercept)	0.00585517	0.076519	
	log(item.pos)	0.00024535	0.015664	0.977
item	(Intercept)	0.00391568	0.062575	
Residual		0.04956476	0.222631	

Number of obs: 696, groups: subject, 29; item, 24

Fixed effects:

	<i>Estimate</i>	<i>Std. Error</i>	<i>t value</i>
(Intercept)	2.010e-01	4.219e-02	4.763
loco1	-7.522e-02	2.395e-02	-3.141
anaphref	-1.531e-02	2.391e-02	-0.640
time.left	1.240e-04	2.554e-05	4.857
log(item.pos)	-3.713e-02	1.054e-02	-3.522
loco1:anaphref	2.036e-03	3.385e-02	0.060

Analysis of Variance Table:

	<i>Df</i>	<i>Sum Sq</i>	<i>Mean Sq</i>	<i>F value</i>	<i>F</i>	<i>Df2</i>	<i>p</i>
loco	1	1.05014	1.05014	21.1872	21.1872	690	0.00000
anaph	1	0.04445	0.04445	0.8968	0.8968	690	0.34397
time.left	1	1.51595	1.51595	30.5853	30.5853	690	0.00000
log(item.pos)	1	0.61885	0.61885	12.4857	12.4857	690	0.00044
loco:anaph	1	0.00018	0.00018	0.0036	0.0036	690	0.95206

Tabelle 7.4.: AnaLoCo 1. Gemischtes lineares Modell für die Residuen der logarithmierten Lesezeiten auf der Position des nachfolgenden Hauptsatzes (Mittelwert der beiden Wörter nach dem finiten Verb des Hauptsatzes).

**Korrektheit der Antworten** Zur Analyse der Korrektheit der Antworten benötigen wir ein Verfahren für binomial verteilte Daten (binäre abhängige Variablen).<sup>23</sup> Wir verwenden dazu ein gemischtes Logit-Modell.<sup>24</sup> In diesem zeigt sich für den Faktor Anapherntyp ein höchst signifikanter Haupteffekt ( $z = 3.374$ ,  $p < 0.001$ ). Entgegen den Erwartungen gibt es aber weder ein Haupteffekt für den Faktor Satztyp noch eine Interaktion zwischen den Experimentalfaktoren (Satztyp:  $z = -0.090$ ,  $p > 0.9$ , Interaktion Satztyp×Anapherntyp:  $z = 0.347$ ,  $p > 0.72$ ).

Generalized linear mixed model fit by the Laplace approximation

Formula: answer ~ loco \* anaph + log(item.pos) + (1 | item) + (1 | subject)

Random effects:

Groups	Name	Variance	Std.Dev.
subject	(Intercept)	0.101931	0.31927
item	(Intercept)	0.014004	0.11834

Number of obs: 684, groups: subject, 29; item, 24

Fixed effects:

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	-0.13481	0.36782	-0.367	0.71398
loco	-0.02028	0.22586	-0.090	0.92847
anaph	0.83539	0.24758	3.374	0.00074 ***
log(item.pos)	0.19101	0.09827	1.944	0.05193 .
loco:anaph	0.12262	0.35300	0.347	0.72831

Tabelle 7.5.: AnaLoCo 1. Logit Mixed Model für die Korrektheit der Antworten.

**Latenz der Beantwortung** Das gemischte Modell für die logarithmierte Latenz der Beantwortung bestätigt wie erwartet den Haupteffekt für den Faktor Anapherntyp (höchst signifikanter Haupteffekt mit  $t = -4.58$ ,  $F = 47.9865$ ,  $p < 0.0001$ ). Für die Latenz zeigt sich aber nun der erwartete Haupteffekt für den Faktor Satztyp als Tendenz ( $t = 1.42$ ,  $F = 3.6623$ ,  $p = 0.052$ , wiederum keine Interaktion Satztyp×Anapherntyp:  $t = 0.10$ ,  $F = 0.0092$ ,  $p > 0.9$ ).

## 7.2.5. Diskussion AnaLoCo 1

Für die Bedingung LSK haben wir einen frühen Effekt der Beschleunigung während der LSK für die Verarbeitung der einfachen und regulären Hauptsatzstruktur der lokalen Subsequenz erwartet. Direkt im Anschluß an die LSK haben wir einen späten Effekt der Verlangsamung für die Zurückweisung der in der LSK angenommenen Satzstruktur auf der nicht integrierbaren Satzposition erwartet. Da sich in den Latenzen der Lesezeiten über den relevanten Satzbereichen auf den frühen Positionen kein Effekt zeigt, kann die

<sup>23</sup>Zur Erinnerung: Die Korrektheit der Antworten wurde binär kodiert mit dem Wert 0 für falsch beantwortete Fragen und dem Wert 1 für richtig beantwortete Fragen.

<sup>24</sup>Gemischte Logit-Modelle gehören zur Klasse der generalisierten linearen Modelle, mit denen auch Daten analysiert werden können, die nicht die Voraussetzung der Normalverteilung erfüllen. Die Funktion *lmer* ermöglicht (unter Verwendung des Parameters *family* = "binomial") das Fitten eines gemischten Logit-Modells zur Analyse binomial verteilte Daten.

Linear mixed model fit by REML

Formula: log.time ~loco \* anaph + answer + log(item.pos) + (1 | item) + (1 + log(item.pos) | subject)

Random effects:

<i>Groups</i>	<i>Name</i>	<i>Variance</i>	<i>Std.Dev.</i>	<i>Corr</i>
subject	(Intercept)	0.0339146	0.184159	
	log(item.pos)	0.0010144	0.031849	-0.034
item	(Intercept)	0.0114052	0.106795	
Residual		0.1156138	0.340020	

Number of obs: 684, groups: subject, 29; item, 24

Fixed effects:

	<i>Estimate</i>	<i>Std. Error</i>	<i>t value</i>
(Intercept)	8.318810	0.072342	114.99
locoloco	0.052783	0.037067	1.42
anaphref	-0.171060	0.037323	-4.58
answer	-0.074799	0.030627	-2.44
log(item.pos)	-0.050296	0.016639	-3.02
locoloco:anaphref	0.005011	0.052283	0.10

Analysis of Variance Table:

	<i>Df</i>	<i>Sum Sq</i>	<i>Mean Sq</i>	<i>F value</i>	<i>F</i>	<i>Df2</i>	<i>p</i>
loco	1	0.4234	0.4234	3.6623	3.6623	678	0.05608
anaph	1	5.5479	5.5479	47.9865	47.9865	678	0.00000
answer	1	0.8258	0.8258	7.1429	7.1429	678	0.00771
log(item.pos)	1	1.0664	1.0664	9.2235	9.2235	678	0.00248
loco:anaph	1	0.0011	0.0011	0.0092	0.0092	678	0.92368

Tabelle 7.6.: AnaLoCo 1. Gemischtes lineares Modell für die Antwortlatenzen.

angenommene Beschleunigung der Lesegeschwindigkeit in unseren Daten nicht bestätigt werden.

Auf den späteren Positionen aber finden sich Effekte. Nach der Präpositionalphrase sollte in der lokalen Hauptsatzstruktur das Satzende oder ein Übergang in einen weiteren Teilsatz folgen. Das tatsächlich folgende finite Verb kann nur in die globale Interpretation integriert werden, nicht aber in die lokal aufgebaute Struktur. Somit ergibt sich eine Deckungsungleichheit der projizierten Wortfolgestruktur, die zu einem höheren kognitiven Aufwand führen sollte. Die tendenziell höhere Lesezeit auf dem Matrixsatzverb entspricht unseren Erwartungen für einen solchen Zusammenbruch der LSK-Interpretation. Allerdings ist der Effekt nicht signifikant, sondern nur als Tendenz vorhanden. Der darauf folgende signifikante Effekt auf dem beginnenden Hauptsatz<sup>25</sup> wurde zwar ebenfalls erwartet, nicht aber die Wirkungsrichtung, die sich in unseren Daten zeigt: Anstatt einer weiteren Verzögerung der Lesegeschwindigkeit im Sinne eines Spillover-Effekts auf der Position der LSK-Auflösung zeigt sich in den Daten eine geringere Latenz für die LSK-Bedingung gegenüber der Kontrollbedingung.

Warum aber sollte unserer Argumentation folgend eine Beschleunigung für die LSK-Bedingung auftreten? Eine sehr spekulative und vage Antwort könnte sein, dass sich dieser Effekt aus dem Zusammenspiel der Eigenschaften der experimentellen Methode und des Zusammenbruchs der lokalen Interpretation ergibt: Während das Ende der LSK eine Reinterpretation der Satzstruktur erfordert, wird eine Reanalyse im linearen Fortschreiten der Satzpräsentation im Self-Paced Reading unterbunden. Entweder gibt der VTn an dieser Stelle das Verstehen des Satzes auf (d.h., er klickt sich nur noch in entsprechend höherer Geschwindigkeit durch das Beispiel durch), oder er sucht nach einer Möglichkeit, einen neuen Ansatz für die Interpretation im Folgesatz aufzubauen. Zugegebenermaßen kann dies keine befriedigende Erklärung sein. Fest steht aber, dass sich die Bedingung LSK auf die Lesezeiten auswirkt.

Die Effekte für die Bedingung Anapherotyp auf den Folgepositionen nach der anaphorischen Referenz zeigen einen deutlichen Vorteil für die Verarbeitung der Reflexivpronomen. Die Unterschiede zeigen sich innerhalb des Satzes, in dem die Interpretation der Anapher relevant ist. Dies dürfte daraus resultieren, dass die Auswahl des Koreferenten für die aufgebaute Argumentstruktur bedeutsam ist, die in diesem Teilsatz aufgebaut wurde. Der Haupteffekt mit dem Vorteil für Reflexivpronomen ist daraus zu erklären, dass die referenzielle Auflösung von Reflexivpronomen in erster Linie durch strukturelle Faktoren bedingt ist, und keine weiteren Constraints erfüllen muss. Folglich kann sie im Unterschied zu Personalpronomen, deren Auflösung die Berücksichtigung weiterer Constraints erfordert,<sup>26</sup> vergleichsweise schnell erfolgen.

<sup>25</sup>Da dieser Effekt in unmittelbarer Folge zur Auflösung der LSK nach dem Matrixsatz-Verb geschieht (mit Überbrückung des topikalisierten Hauptsatzverbs, für das sich überhaupt kein Effekt einstellt), ist es absolut schlüssig, diese Position im Sinne eines verzögerten Effekts als für unsere Fragestellung relevant zu erachten und in die Analyse mit aufzunehmen. Bei der Stärke des gefundenen Effekts sind wir dazu gezwungen, ihn zu diskutieren und in Beziehung zu den Experimentalbedingungen zu setzen.

<sup>26</sup>Für Personalpronomen können noch andere Referenten in Frage kommen, die nicht einmal innerhalb desselben Satzes genannt sein müssen. Zudem ist eine kataphorische Verwendung des Personalpronomens möglich.

In den Daten ist kein Hinweis auf eine Interaktion – welcher Art auch immer – zu finden. So müssen wir zunächst davon ausgehen, dass sich LSK nicht unterschiedlich auf verschiedene Anapherentypen auswirken. Dies ist verwunderlich, da wir erstens deutliche Unterschiede der Verarbeitung von Reflexiv- und Personalpronomen in Form von Haupteffekten gefunden haben. Zweitens haben wir für LSK angenommen, dass sie sich vor allem auf der Ebene syntaktisch-struktureller Constraints und deshalb stärker oder früher auf Reflexivpronomen auswirken. Denkbar wäre, dass sich LSK hauptsächlich auf einen primär strukturell operierenden initialen Filter auswirken, und die gefundenen LSK-Effekte deshalb beide anaphorischen Formen gleichermaßen beeinflussen.

Für die Korrektheit der Antworten zu den Fragen wurden keine Effekte der LSK gefunden. Diese scheinen keine längerfristige Auswirkung auf die Interpretation des Satzes zu haben. Da wir davon ausgehen, dass der LSK-Effekt nur ein temporäres Phänomen während der Satzverarbeitung ist, das nur lokal die Interpretation beeinflusst, ist das Ergebnis nicht allzu überraschend. Erstaunlich hingegen ist der Befund einer starken Tendenz für den Unterschied der Satztypen in den Latenzen zur Beantwortung der Fragen. Dieser Befund entspricht unseren Erwartungen mit einer geringeren Latenz für die Kontrollbedingung und mit einer stärkeren Verzögerung für die LSK-Bedingung – LSK scheinen zu einer verlangsamten Reaktion bei der Beantwortung der Fragen zu führen. Das Ausbleiben einer Interaktion mit dem Faktor Anapherentyp bestätigt die Befunde zu den Lesezeiten – LSK scheinen sich auf beide Anapherentypen in gleicher Weise auszuwirken.

Für den Faktor Anapherentyp zeigen sich starke Effekte sowohl für die Korrektheit der Antwort als auch für die Latenz der Beantwortung. Diese weisen für die Bedingung mit Reflexivpronomen auf eine wesentlich schnellere Beantwortung der Fragen und einen wesentlich höheren Anteil korrekt beantworteter Fragen hin. Um die Ursache für diesen Effekt zu finden, muss die Struktur der zugrundeliegenden Frage in Betracht gezogen werden. Mit diesen wird die korrekte Zuweisung des Agens und des Patiens zu den beiden Verben im vorangestellten Nebensatz und im Relativsatz geprüft. Damit kann indirekt festgestellt werden, ob der Anapher das jeweils korrekte Antezedens zugewiesen wurde. Während sich dies im Fall der Reflexivpronomen strukturell eindeutig bestimmen lässt, dienen strukturelle Constraints bei Personalpronomen nur im Sinne einer Einschränkung, während zur eindeutigen Bestimmung weitere Constraints (in stärkerer Gewichtung) erwartet werden. Damit ist bereits die Ausgangslage zur Auflösung der beiden Anapherentypen unterschiedlich. Eine alternative Erklärung zum Nachteil der Personalpronomen wäre, dass auf der Anaphernposition im Fall eines Verweises auf den Agens des vorausgegangenen Relativsatzes eine andere Proform erwartet werden würde. Denkbar ist zum Beispiel, dass beim Verweis auf einen direkt zuvor eingeführten Referenten ein Demonstrativpronomen (*dieser*) erwartet oder zumindest bevorzugt verwendet wird. Dies würde bedeuten, dass die Verwendung des Personalpronomens aufgrund der Möglichkeit, auf dieser Satzposition eine Alternativform zu verwenden, dispräferiert ist. Deshalb wäre ein entsprechend starker Effekt bei der Beantwortungsaufgabe zu erwarten.

Zusammenfassend können wir feststellen, dass sich sowohl LSK als auch die Wahl des Anapherentyps auf die Lesezeiten und auf Merkmale der Beantwortung anschließender Fragen auswirken. Allerdings haben sich die Effekte weitgehend nicht in der erwarteten Form gefunden. Zudem mussten auch viele Fragen darüber unbeantwortet bleiben, auf welche Art sich die Faktoren auswirken. Mit der Methode der Lesezeitmessung mit Self-Paced

Reading sind die Möglichkeiten jedenfalls sehr beschränkt, dieses Phänomen zu erforschen. Da wir hieraus nur indirekt auf die Interpretation des Satzes schließen können, lassen sich keine weitergehenden Aussagen zu Präferenzen bei der Antezedens-Auswahl einer Anapher treffen. Denkbar ist zum Beispiel, dass wir nur deshalb keine Interaktionen gefunden haben, weil sich in den Lesezeiten nur frühe strukturelle Effekte der Anapherauflösung widerspiegeln, im Sinne von Auswirkungen eines initialen Filters. Denkbar ist auch, dass LSK ein ausschließlich syntaktisches Phänomen sind – d.h., dass sie die Interpretation der Argumentstruktur der darin erscheinenden Verben beeinflussen, sich aber überhaupt nicht auf die Auflösung anaphorischer Referenzen auswirken. Aus all diesen Feststellungen schließen wir, dass die Erforschung des Phänomens und seiner tatsächlichen Auswirkung die Verwendung exakterer Instrumente erfordert. Die bisherigen Befunde sind vielversprechend genug, um eine weitergehende Untersuchung zu rechtfertigen. Die folgenden Experimente sollen über die Messung von Blickbewegungen einen besseren Einblick in das Phänomen ermöglichen.

### 7.3. AnaLoCo 2: Visual-World-Experiment mit Referenten

Die Hinweise aus der ersten Studie deuten auf einen Einfluss von LSK auf die Interpretation anaphorischer Referenzen hin, lassen aber auch viele Fragen offen. In den folgenden Studien wollen wir herausfinden, ob es sich dabei um ein Oberflächenphänomen oder um einen tatsächlichen Einfluss auf die Interpretation handelt – eine Beeinflussung der koreferenziellen Anbindung der Anapher. Um einen Einblick in die tatsächlichen Bindungspräferenzen an den kritischen Positionen zu erhalten, wurde Experiment 2 als Eyetracking-Studie im Visual-World-Paradigma durchgeführt. Das Visual-World Paradigma wurde mehrfach erfolgreich in unseren Untersuchungsdomänen eingesetzt – zur Untersuchung sowohl des Einflusses von LSK (u.a. Konieczny, Müller, Hachmann et al., 2009; Konieczny, Müller, Baumann et al., 2009; Konieczny & Müller, 2010, , vgl. Kap. 5.2) als auch der Verletzung von Bindungsdomänen (u.a. Runner et al., 2003, 2006; Kaiser et al., 2009, vgl. Kap. 5.3). Im folgenden Experiment werden die erwarteten Effekte lokaler syntaktischer Kohärenzen und die erwarteten Effekte der Online-Anapherauflösung im Visual-World-Paradigma kombiniert. Die hier vorgestellten Ergebnisse wurden in verschiedenen Phasen des Projekts von Weldle, Konieczny, Wolfer et al. (2010), Konieczny, Weldle et al. (2009) und Konieczny, Weldle, Wolfer et al. (2010) präsentiert.

Bei Visual-World-Experimenten findet sich eine starke Verschränkung zwischen der visuellen Aufmerksamkeit und sprachlich beschriebenen Aspekten der Szene bzw. benannten Referenten in der Szene (vgl. Kap. 5.2.2). Wir gehen dabei von der Annahme aus, dass demnach bei der Verarbeitung eines Reflexiv- oder Personalpronomens die visuelle Aufmerksamkeit auf denjenigen Referenten gerichtet ist, der das angenommene Antezedens des Pronomens ist. Die lokale Kohärenz bietet in unseren Zielsätzen ein alternatives Antezedens, das in der globalen Lesart unzugänglich ist. Wenn lokale Kohärenzen tatsächlich verarbeitet werden, sollte zumindest kurzzeitig das alternative Antezedens an das Pronomen gebunden werden. Hierfür wird in den Blickbewegungsdaten eine vermehrte Fixation des Distraktors (lokaler vs. global korrekter Referent) während des kritischen Zeitraums (Nennung des Personal- oder Reflexivpronomens und Folgeregion) erwartet.

### 7.3.1. Design

Wir verwenden wiederum ein zum Self-Paced Reading-Experiment identisches 2×2-Design mit den Faktoren Anapherentyp und Satzstruktur. Der Faktor Anapherentyp erhält die beiden Ausprägungen Reflexivpronomen und Personalpronomen. Der Faktor Satzstruktur erscheint in den Ausprägungen Struktur mit lokaler Kohärenz und Kontrollstruktur mit Unterbindung der lokalen Kohärenz. Die Satzstruktur ist ebenfalls identisch zu den Materialien im ersten Experiment, die Unterbindung der LSK in Faktor 2 wird also dementsprechend durch ein eingefügtes Adverb realisiert. Allerdings entfällt in Experiment 2 im Unterschied zum ersten Experiment die angeschlossene Frage, und damit der Pseudo-Faktor zur Korrektheit der beantworteten Frage. Es bleibt hier also bei einem reinen 2×2-Design (zur Kodierung der Bedingungen vgl. Tab. 7.7).

Design		Faktor 1: Anapherentyp	
		sich	ihn
Faktor 2: LSK	mit LSK	1	2
	LSK durchbrochen	3	4

Tabelle 7.7.: Tabelle zum Design für Experiment 2.

**Bedingung 1** Reflexivpronomen (*sich*) mit lokal interferierender Lesart (pro LSK).

**Bedingung 2** Personalpronomen (*ihn*) mit lokal interferierender Lesart (pro LSK).

**Bedingung 3** Reflexivpronomen (*sich*), unterdrückte lokale Lesart (contra LSK).

**Bedingung 4** Personalpronomen (*ihn*), unterdrückte lokale Lesart (contra LSK).

### 7.3.2. Stimulus-Materialien

**Sprachlich-auditive Stimuli** Das sprachliche Stimulusmaterial entspricht in der Struktur genau dem Material, das im Lesezeiten-Experiment eingesetzt wurde: Ein vorangestellter temporaler Nebensatz mit eingebettetem Relativsatz, gefolgt von einem Hauptsatz. Dabei schließt sich direkt an den Relativsatz das Matrixsatz-Objekt in Form einer anaphorischen Referenz (Reflexiv- oder Personalpronomen) an. Das Satzmaterial enthält also eine lokale Kohärenz, wobei die kritische Position innerhalb der lokalen Kohärenz eine anaphorische Referenz enthält. Global verweist die Anapher im reflexiven Fall eindeutig auf einen Referenten. Im transitiven Fall hingegen schließt sie diesen eindeutig aus: Das Reflexivpronomen *sich* bezieht sich in der globalen Lesart zwingend auf das Subjekt des Matrixsatzes (temporaler Nebensatz). Das Personalpronomen *ihn* steht immer im Genus Maskulinum und im Kasus Akkusativ. Es schließt komplementär zum Reflexivpronomen das Subjekt des Matrixsatzes obligatorisch aus. Global mögliche Antezedenzen sind entweder das Subjekt des Objekt-Relativsatzes oder ein beliebig anderes als Maskulinum Singular markiertes Antezedens.

In der LSK-Bedingung mit Objekt-Relativsatz wird über die Relativsatzgrenze hinweg eine LSK aufgebaut, die das Pronomen mit einschließt. Mit dem Subjekt des Relativsatzes wird ein lokales Antezedens eingeführt. Für die Kontrollbedingungen haben wir wiederum

Adverbien verwendet, die das Relativsatzverb modifizieren und damit die Ausdehnung der lokalen Kohärenz über die Relativsatzgrenze hinweg unterbinden. Beispiel 27 verdeutlicht diese Struktur in den vier experimentellen Bedingungen: Reflexivpronomen in der LSK (Bsp. 27b), Personalpronomen in der LSK (Bsp. 27c). Kontrollversionen: Reflexivpronomen in durchbrochener LSK durch Adverbeinschub (Bsp. 27d), Personalpronomen in durchbrochener LSK durch Adverbeinschub (Bsp. 27e).

- (27) a. In der folgenden Szene sehen wir einen Vater, einen Sohn und einen Clown.  
b. Als der Vater<sub>i</sub>, den der Sohn<sub>j</sub> kämmt, sich<sub>i/\*j</sub> im Wohnzimmer anzieht, ist die halbe Stadt bereits auf dem Weg zur Aufführung.  
c. Als der Vater<sub>i</sub>, den der Sohn<sub>j</sub> kämmt, ihn<sub>\*i/j...</sub> im Wohnzimmer anzieht, ist die halbe Stadt bereits auf dem Weg zur Aufführung.  
d. Als der Vater<sub>i</sub>, den der Sohn<sub>j</sub> gerade kämmt, sich<sub>i/\*j</sub> im Wohnzimmer anzieht, ist die halbe Stadt bereits auf dem Weg zur Aufführung.  
e. Als der Vater<sub>i</sub>, den der Sohn<sub>j</sub> gerade kämmt, ihn<sub>\*i/j...</sub> im Wohnzimmer anzieht, ist die halbe Stadt bereits auf dem Weg zur Aufführung.

Für die Verwendung im Visual-World-Paradigma haben wir dieses Satzmaterial akustisch aufbereitet und dazu das Satzmaterial in normalem Sprechtempo gesprochen.<sup>27</sup> Beim Einsprechen der Sätze wurde darauf geachtet, dass diese möglichst neutral gesprochen wurden, d.h. ohne allzu auffällige Intonation. Anschließend wurde es mit einem Audiotbearbeitungsprogramm bearbeitet.<sup>28</sup> Die Bearbeitungsschritte werden in den folgenden Abschnitten kurz erläutert. Die Methodik folgt dabei der Vorgehensweise, die unter anderem von Konieczny, Müller, Hachmann et al. (2009) verwendet wurde.

Die LSK in unseren Ziel-Sätzen überschreiten die Satzgrenze des Relativsatzes. Diese Satzgrenze ist auch prosodisch durch Pause und Intonation markiert. Prosodie ist ein relativ starker Hinweisreiz für den Aufbau der syntaktischen Struktur (vgl. die Diskussion unter Kap. 5.2.3). In unserem Experiment wollen wir aber rein strukturell-syntaktische Effekte untersuchen. Deshalb mussten die prosodischen Marker, die auf das Ende des Relativsatzes hinweisen, neutralisiert werden. Durch diese Neutralisierung nicht-syntaktischer Hinweisreize für die korrekte syntaktische Anbindung erhält das Material einen Bias zu Gunsten der Verarbeitung der LSK: Die Grenze des Relativsatzes kann nicht mehr so gut erkannt und die LSK-Lesart folglich nicht so gut ausgeschlossen werden. Ziel war es, die korrekte syntaktische Abfolge und Segmentierung des Satzes zu erhalten, während Strukturhinweise aus Prosodie und Intonation zerstört werden. Als günstiger Nebeneffekt ergibt sich, dass wir für alle vier Bedingungen den gleichen Ausgangssatz verwenden können. Dies verbessert auch die allgemeine Kontrolle über die sprachlichen Stimuli, da individuelle Variationen über die Sätze hinweg neutralisiert werden. Eine kritische Diskussion dieser Verfahrensweise folgt an späterer Stelle (vgl. Abschnitt 7.3.6).

Um die Effekte der prosodischen Hinweisreize für die Relativsatzgrenze zu neutralisieren, haben wir das im Folgenden beschriebene Tonspur-Einschneide-Verfahren eingesetzt.

---

<sup>27</sup>Das Satzmaterial wurde einheitlich von nur einer Person gesprochen. Vielen Dank an dieser Stelle an Sarah Schwarzkopf.

<sup>28</sup>Wir haben hierfür das Open-Source-Programm Audacity gewählt, das frei zugänglich ist und eine relativ einfache und komfortable Bearbeitung ermöglicht. Audacity-Download auf der Projektseite unter: <http://audacity.sourceforge.net>

Zunächst wurde zusätzlich zu den Experimentalsätzen für jeden Satz eine Version der LSK ausgesprochen. Die Sequenz wurde als Hauptsatz mit Fortsetzung gesprochen. In unserem Beispiel entspricht dies dem folgenden Satz: *Der Sohn kämmt sich im Wohnzimmer der Großmutter*. Die Fortsetzung dient dazu, einer Intonation des Satzendes vorzubeugen. Anschließend wurde diese isolierte Sequenz so gekürzt, dass nur noch die gewünschte LSK gegeben ist. Dabei wurde auch der Artikel am Satzanfang entfernt, um die Satz-anfangsintonation zu eliminieren. Diese isolierte Sequenz haben wir zu vier Versionen weiterverarbeitet. Hierfür wurden zuerst aus dem primären Satzmaterial das Personalpronomen *ihn* und aus jedem Basissatz das Adverb für die Kontrollsätze (z.B. *gerade*) herausgeschnitten. Anschließend wurden aus der LSK vier Versionen erstellt: Version (a) bleibt wie gehabt, für Version (b) wurde an der Stelle des Reflexivpronomens das Personalpronomen eingeschnitten, für Version (c) wurde vor das finite Verb das isolierte Adverb eingeschnitten, und für Version (d) wurde sowohl das Adverb als auch das Personalpronomen eingeschnitten. Außer den kritischen Manipulationen sind die Sätze also identisch. Im Basissatz, d.h. dem kompletten Satz mit natürlicher Prosodie, wurden die Sprechpausen an den verschiedenen Satzgrenzen bearbeitet, um ein gleichmäßigeres und flüssigeres Sprachbild zu erhalten. Anschließend wurden in diesen Basissatz die vier Versionen der LSK eingeschnitten, um die vier Bedingungen des Designs zu erhalten.

Für die spätere Auswertung der Daten haben wir vorab für jeden Satz die folgenden Zeitmaße aufgezeichnet: (Z.a) Die Gesamtlänge des jeweiligen Satzes, (Z.b) der On-Set des Pronomens, (Z.c) der Off-Set der lokalen Kohärenz. Am Beispiel: *Während der Vater, den der Sohn kämmt, (Z.b) sich zur selben Zeit (Z.c) anzog, war bereits die halbe Stadt auf dem Weg zur Aufführung (Z.a)*. Diese Zeitmarken wurden für die spätere Analyse zur Synchronisation der kritischen Positionen verwendet.

Außer den Zielsätzen wurde auch noch für jeden Satz ein Einleitungssatz ausgesprochen, der die im Satz auftretenden Personen benennt, sowie eine weitere Person, die nicht explizit im Satz auftritt. Dies wird an späterer Stelle erläutert (vgl. Bsp. 27a). Die Reihenfolge der Benennung wurde über die Sätze hinweg durchrotiert, so dass über die Benennungsreihenfolge nicht auf die Erscheinungsreihenfolge im folgenden Ziel-Satz geschlossen werden konnte und auch eventuelle Salienzeffekte durch die Reihenfolge ausgeglichen werden.

Zusätzlich zum Satz-Material aus dem Lesezeiten-Experiment wurden 24 weitere Sätze erstellt, d.h. es wurden nach dem erläuterten Schema insgesamt 48 Experimentalsätze in den vier Bedingungen konstruiert. Dies ergibt ein Gesamtmaterial von 192 Sätzen.<sup>29</sup> Die Bedingungen für die 48 Experimental-Items wurden anhand der Latin-Square-Methode vollständig durchrotiert. Das Experiment wurde parallel zu einem weiteren Visual-World-Experiment durchgeführt, dessen Items als Füllmaterial dienten. Dieses umfasst ebenfalls 48 Experimentalsätze in vier Bedingungen und weist eine zu unseren Experimental-Items vergleichbare Komplexität auf. Zudem wurde in den beiden Studien dasselbe visuelle Präsentationsschema für die Darbietung der visuellen Materialien verwendet (vgl. folgender Abschnitt). Insgesamt ergibt sich daraus ein Verhältnis von 1:1 (50%) zwischen Experimental-Items und Füll-Items.

---

<sup>29</sup>Der elektronische Anhang im PDF-Format enthält das vollständige Stimulus-Material des Experiments AnaLoCo 2 – das heißt, sowohl die sprachlichen als auch die visuellen Stimuli. Der Anhang wird auf Anfrage (per E-Mail an [helmut.weldle@googlemail.com](mailto:helmut.weldle@googlemail.com)) zugesandt.

**Visuelle Stimuli** Für die visuellen Stimuli verwenden wir im Unterschied zu den Studien von Konieczny, Müller, Hachmann et al. (2009) keine szenischen Darstellungen, sondern Abbildungen von Einzelpersonen. In den zugrundeliegenden Sätzen haben wir hierfür immer Personenbezeichnungen benutzt, die sich gut zeichnerisch darstellen und hinsichtlich ihrer Merkmale leicht unterscheiden lassen. Dies schließt vor allem die Klasse der Berufsbezeichnungen ein. So kamen unter anderem folgende Personen zum Einsatz: Polizist, Bauer, Clown, Soldat, Jongleur, Demonstrant, Arzt, Lehrer, Bergsteiger, Feuerwehrmann, Bankdirektor, etc. Diese Personen wurden in einfacher Skizzenform als Schwarz-Weiß-Zeichnungen abgebildet.<sup>30</sup>

Für jeden Satz wurde ein korrespondierendes Set von drei Personen-Abbildungen zusammengestellt (vgl. 7.6): Abbildung (a) entspricht dem Subjekt des Matrixsatzes (des vorangestellten Nebensatzes) – in unserem Satzbeispiel 27 ist dies der VATER. Abbildung (b) entspricht dem Subjekt des Relativsatzes – in unserem Satzbeispiel 27 der SOHN. Als dritte Abbildung (c) haben wir eine Person hinzugefügt, die in den Kontext der beschriebenen Szene passt bzw. in losem Zusammenhang zu diesem Kontext steht, vor allem in Bezug zum nachfolgenden Hauptsatz. Die Person selbst wird aber im Satz nicht namentlich erwähnt. Dies war in unserem Satzbeispiel 27 ein CLOWN, da im folgenden Hauptsatz eine Zirkusaufführung erwähnt wird, die die zuvor genannten Personen besuchen wollen.

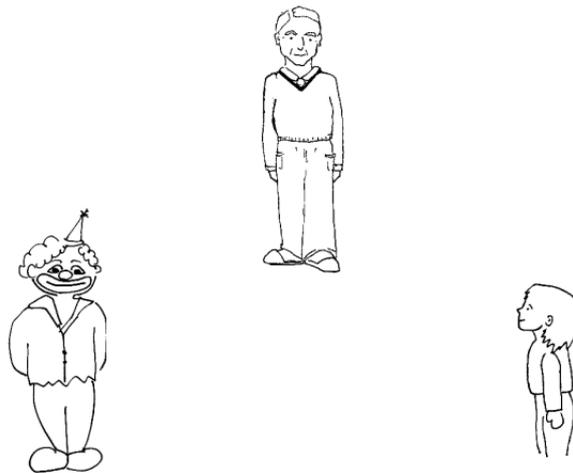


Abbildung 7.6.: Visuelle Stimuli im Visual-World-Experiment AnaLoCo 2: Abbildungen des Vaters und des Sohns korrespondierend zur Benennung im zugehörigen Ziel-Satz sowie eines nicht explizit genannten Clowns.

Bei der Zusammenstellung der jeweiligen Personen für eine Szene haben wir darauf geachtet, dass sich diese visuell deutlich voneinander unterscheiden und somit einfach zu erkennen und den sprachlichen Benennungen zuzuordnen sind. Um auszuschließen, dass die VTn während des Hörens des Ziel-Satzes erst die abgebildeten Personen den verbalen Benennungen zuordnen müssen, wurden die drei vorkommenden Personen jeweils im oben erwähnten vorgeschalteten Einleitungssatz benannt (vgl. Bsp. 27a). Die Einzelpersonen

<sup>30</sup>Für die Zeichnungen wurde eine Grafik-Design-Studentin eingestellt, die die Personen nach unseren Vorgaben erstellt hat.

wurden in der Größe 650 x 450 Pixel abgebildet (bei einer Gesamtgröße des Bildschirms von 20 Zoll und einer Auflösung von 1600 x 1200 Pixel). Die Positionierung erfolgte als Anordnung in einem gleichseitigen Dreieck mit einer mittig-oben platzierten Abbildung und zwei unterhalb nebeneinander gesetzten Abbildungen. Die Positionsauswahl für die Abbildungen wurde über die Bedingungen hinweg nach der Latin-Square-Methode rotiert, unabhängig von der Rotation der sprachlichen Stimuli. Jeder Versuchsteilnehmer erhält damit jeden Satz nur in einer Bedingung, wobei die dazugehörige Bilderkonstellation durch ihre Rotation weder zu einer Bedingung noch zur Reihenfolge der Personenbenennung im Zielsatz korrespondiert.

### 7.3.3. Methodik und Hypothesen

Das Vorgehen entspricht der Methodik des Visual-World-Paradigmas und folgt im Prinzip dem experimentellen Set-Up von Konieczny, Müller, Hachmann et al. (2009). Eher untypisch ist die Verwendung von handlungsunabhängigen Einzelpersonen anstelle einer oder mehrerer zusammenhängender Szenen. Die Grundidee folgt dabei dem Vorbild der Studien von Runner et al. (2006, 2003). Zur Aufzeichnung der Fixationen wurden nicht-überlappende IAs definiert, die in ihrer Ausdehnung der Größe der Einzelpersonen-Abbildungen entsprechen. Die hierauf gerichteten Fixationen wurden in ihrer zeitlichen Auflösung auf die Zeitsignatur des dargebotenen Satzes abgebildet.

Zur Implementierung des Experiments wurde die Software Experiment Builder von SR Research verwendet. Die Durchführung und Aufzeichnung erfolgte über ein Remote-Eyetracker-System (ohne Verwendung einer Kopfstütze). In den kognitionswissenschaftlichen Forschungslabors stand dafür das System Eyelink 1000 (ebenfalls SR Research) zur Verfügung.

#### 7.3.3.1. Ablauf und Teilnehmer

Am Experiment nahmen 25 VTn teil, vorwiegend Studierende der Universität Freiburg. Die Teilnehmer erhielten im Anschluss an das Experiment einen Versuchsteilnehmerstundenschein oder Geld.<sup>31</sup> Vor dem Experiment wurden in Papierform Instruktionen zum Ablauf und zur Aufgabe gegeben. Nach dem Lesen der Instruktionen konnten die VTn Verständnisfragen stellen. Das Experiment wurde parallel zu einem weiteren Visual-World-Experiment durchgeführt, dessen Items wie oben beschrieben als Füllmaterial für das Experiment dienten. Es wurde als Teil einer größeren Serie an Experimenten durchgeführt, die die VTn in zwei Sitzungen ableisteten.<sup>32</sup>

---

<sup>31</sup>Die Vergütung erfolgte wieder in Abhängigkeit von der benötigten Zeit und betrug 5,- Euro pro 30 min.

<sup>32</sup>Dazu gehörten ein Anomaly Detection Experiment sowie zwei Tests zur sprachlichen Leistungsfähigkeit (Bestandteil der Magisterarbeit von Hachmann, 2008; Hachmann et al., 2009): Ein Lesespannentest (in einer Version nach Hacker, Handrick & Veres, 1996) und ein Test mit verbaler Sortieraufgabe (Mendelsohn, 2002, theoretisch und methodisch angelehnt an den Wisconsin Card Sorting Task). Zwischen den Experimenten mussten die VTn zudem einen Fragebogen zu ihren Lesegewohnheiten und ihrem Sprach- und Schriftspracherwerb ausfüllen. Ziel der Experimentalserie war die Untersuchung von Zusammenhängen zwischen der Empfänglichkeit für die Interpretation lokaler Kohärenzen und interindividuellen Differenzen bei der Sprachverarbeitung. Kurz gefasst: Sind VTn mit schlechterer Verstehensleistung empfänglicher für LSK-Effekte?

Die Aufgabe der VTn war es, sich die Sätze anzuhören und parallel dazu auf dem Bildschirm die statischen Abbildungen der Referenten zu betrachten. Nach zwei Übungsdurchgängen folgten die Experimental-Items. Zum Starten der Items mussten die VTn jeweils eine Tastaturtaste drücken (hier: Leertaste). Hierdurch wurde die jeweilige Abbildung mit allen drei Personen eingeblendet, die bis zum Abschluss des Trials permanent sichtbar blieb. Das sprachliche Stimulusmaterial wurde per Lautsprecher auditiv präsentiert. Nach dem beschreibenden Einleitungssatz folgte der jeweilige Zielsatz. Die Primär-Aufgabe der VTn war es, das gesprochene Material anzuhören und währenddessen die Abbildungen zu betrachten. Die Sekundär-Aufgabe war es, nach Abschluss des Zielsatzes (markiert durch einen Glockenton unmittelbar nach Ende des Satzes) mit der Computermaus diejenige Person anzuklicken, die ihrer Meinung nach in den beschriebenen Ereignissen die wichtigste Rolle spielt. Für diese Sekundär-Aufgabe gab es keine Rückmeldung.<sup>33</sup> Die zusätzliche Aufgabe sollte sicherstellen, dass die VTn dem gesprochenen Material inhaltlich folgen und einen Bezug zwischen dem gehörten Satz und den abgebildeten Personen herstellen.

#### 7.3.3.2. Hypothesen

Der Verwendung der Eyetracking-Methode liegt die Annahme zugrunde, dass bei der Abbildung von mehreren vorhandenen Referenten die visuelle Aufmerksamkeit auf den Referenten gerichtet ist, auf den aktuell sprachlich referiert wird. Bei der Verarbeitung eines Reflexiv- oder Personalpronomens sollte die visuelle Aufmerksamkeit demnach auf denjenigen Referenten gerichtet sein, der das angenommene Antezedens des Pronomens ist. Somit sollte sich anhand dieser Methode mit hoher zeitlicher Auflösung messen lassen, auf welcher Satzposition welches Antezedens der Interpretation zugrundegelegt wird. Unseren Hypothesen folgend sollte bei der Anbindung von Reflexiv- oder Personalpronomen, denen durch ihre Einbettung in eine LSK eine alternative Anbindungsmöglichkeit gegeben ist, diese lokale Bindungsdomäne bei der Auswahl bzw. beim Ausschluss des Antezedens in Betracht gezogen werden. Dies sollte sich in den Blickbewegungsdaten in der Version mit LSK in einem höheren Anteil an Fixationen auf dem global unzugänglichen, aber temporär im lokalen Kontext konkurrierenden Antezedens bemerkbar machen.

Hinsichtlich des genauen Einsatzes der Effekte eine Hypothese zu formulieren ist schwierig, da es keine direkt vergleichbaren Daten gibt. In verschiedenen Befunden werden frühe und späte Effekte verschiedener Einflussfaktoren auf die Anaphernauflösung festgestellt (vgl. z.B. Sturt, 2003; Kaiser et al., 2009; van Gompel & Majid, 2004; Badecker & Straub, 2002). Speziell für strukturelle Effekte wurde übereinstimmend ein früher Einsatz gefunden. Es ist allerdings schwer, daraus einen Zeitraum abzuleiten, in dem wir Effekte für unser Design erwarten würden. Zunächst einmal beruhen die Befunde auf unterschiedlichen experimentellen Methoden, die häufig genug keine näheren Angaben zur genauen zeitlichen Auflösung bieten können (z.B. Self-Paced Reading bei Badecker & Straub, 2002; Kennison, 2003). Die meisten Studien wurden in englischer Sprache durchgeführt. Ob sich die Daten exakt mit denen einer deutschsprachigen Studie decken werden und genau die gleichen strukturellen Effekte erwartet werden dürfen, können wir vorab nicht beantworten. Zeitliche Effekte werden immer auch in Relation zu bestimmten induzierten

---

<sup>33</sup>Wir konnten hierfür keine Rückmeldung einblenden, da es faktisch keine korrekte Antwort gibt. Dennoch wurde die Antwort der Sekundär-Aufgabe (angeklickte Person) aufgezeichnet.

Effekten verschiedener Einflussfaktoren genannt. Inwiefern wir daraus Vorhersagen für die in unserem Design kontrollierten Einflussfaktoren ableiten können, bleibt ebenfalls nicht beantwortbar, denn die Klasse dieser Faktoren unterscheidet sich deutlich von den bisher untersuchten. Die Effekte der Studie von (Kaiser et al., 2009), in der ebenfalls Reflexiv- und Personalpronomen untersucht werden, finden sich in einem Zeitraum zwischen 200 und 1800 ms. Um die Aussagen aus Kapitel 5.3.2 grob vereinfacht zusammenzufassen: Generell scheinen Reflexivpronomen stärker durch strukturelle Faktoren bestimmt zu werden als Personalpronomen, auf die sich semantische Faktoren stärker auswirken können. Allerdings zeigen sich auch für Personalpronomen frühe strukturelle Effekte. Da wir in unserem Design keine semantischen Einflussfaktoren kontrollieren und die Effekte auf rein strukturelle Faktoren (hier: LSK) zurückzuführen sind, erwarten wir eine stärkere Wirkung der LSK auf Reflexivpronomen. Ob die Effekte deshalb aber bei Reflexivpronomen früher einsetzen und länger anhalten als bei Personalpronomen, bleibt spekulativ.

Im Detail erwarten wir für die beiden Anapherarten folgende Effekte, die sich auf den zeitlichen Verlaufs ihrer Verarbeitung, auf das Einwirken der experimentellen Bedingungen sowie auf die Verschiebung der Anteile der visuellen Aufmerksamkeit beziehen:

### 1. Auswirkungen und Einsatz der Effekte für das Reflexivpronomen.

- Wir erwarten eine Verschiebung der relativen Häufigkeit der Fixationen vom Agens des Matrixsatzes (Subjekt des temporalen Nebensatzes) auf das lokale Agens (Subjekt des Relativsatzes). Im Beispielsatz entspricht dies der Verschiebung der Fixationswahrscheinlichkeit auf der Abbildung des VATERS hin zur Abbildung des SOHNS. Dies entspricht einer Lesart entgegen des globalen strukturellen Bindungsprinzips A.
- Folglich erwarten wir in der LSK-Bedingung (Bedingung 1 im Vergleich zu Bedingung 3) eine Steigerung des Gesamtanteils der Fixationen auf dem lokalen Antezedens entsprechend einer lokal isolierten Anwendung des Prinzips A. Diese müsste einhergehen mit einer Verringerung des Gesamtanteils der Fixationen in der LSK-Bedingung auf dem global korrekten Antezedens.
- Da wir ausschließlich strukturelle, aber keine semantischen Einflussfaktoren variieren, und strukturelle Bindungsprinzipien als früherer Filter wirken, sollten sich die Effekte zu einem frühen Zeitpunkt zeigen. Denkbar ist, dass sich die Auflösung der Referenz in der ambigen Anbindungsituation zwischen lokaler und globaler Lesart im Fall der Reflexivpronomen stärker an strukturellen Faktoren orientiert und sich folglich für diese ein früherer Zeitpunkt der Reorientierung (LSK-Effekt) zeigt als für Personalpronomen.<sup>34</sup>

### 2. Auswirkungen und Einsatz der Effekte für das Personalpronomen.

- Die Erwartungen der Effekte auf die Fixationen sind im Fall der Personalpronomen etwas komplexer, da die hier angenommenen strukturellen Constraints nicht determinierend (Auswahl eines einzigen Agens), sondern nur beschrän-

---

<sup>34</sup>Um eine Schätzung für einen frühen Effekt zu geben: Mit einer Annäherung an die Zeit, die für das vollständige Aussprechen der Anapher (ca. 250 ms) sowie für die Sakkadenplanung (ca. 250 ms) benötigt wird, sollten wir ab ca. 500 ms einen Effekt finden.

kend (Ausschluss eines Agens, aber Aufrechterhaltung aller grammatikalisch und semantisch kompatiblen Kandidaten) wirken. Nach Prinzip B ist das Subjekt des Matrixsatzes für die Bindung unzugänglich. Das Personalpronomen kann also nur mit dem Subjekt des Relativsatzes oder einem beliebigen anderen grammatisch kongruenten Referenten koreferieren. In der LSK-Bedingung sollte der relative Anteil der Fixationen auf dem Subjekt des Relativsatzes aber reduziert sein, da es im lokalen Kontext die Rolle des c-kommandierenden Arguments einnimmt, und deshalb entsprechend Prinzip B strukturell ausgeschlossen ist.

- Der Betrachter muss nun eine Alternative zu diesem Referenten finden, der durch seine explizite Benennung im Satz sehr präsent ist. Hierfür kommt entweder das global unzugängliche Subjekt des Matrixsatzes in Betracht, das in der LSK keine c-kommandierende Position belegt, oder aber der dritte unbeteiligte Referent, der allerdings aufgrund der fehlenden Benennung an dieser Stelle nicht sehr salient sein dürfte. Im Beispielsatz entsprechen diese Auswirkungen einer Verschiebung der Fixationen von der Abbildung des SOHNS auf die des VATERS und eventuell des CLOWNS.
- Zwar sollten sich die strukturellen Einflussfaktoren gleichermaßen als früher Filter auf die Personalpronomen auswirken. Denkbar ist aber auch, dass sich die Auflösung der Referenz in der ambigen Anbindungssituation eher an nicht-strukturellen Constraints orientiert und sich deshalb eine Verzögerung des Effekts der strukturellen Einflussfaktoren ergibt – d.h. der Effekt zeigt sich für Personalpronomen später als für Reflexivpronomen. Andererseits sollte sich die Reduktion der Fixationen auf dem Subjekt der LSK (auf dem SOHN) unmittelbar einstellen und ebenfalls zu einem frühen Effekt führen, da hier die Ausschlussfunktion der früh wirkenden strukturellen Constraints einsetzen müsste.

#### 7.3.4. Ergebnisse

Die Methodik der Auswertung richtet sich weitgehend nach den Analysen für Visual-World-Daten, wie sie von Konieczny und Müller (2007); Konieczny, Müller, Hachmann et al. (2009) durchgeführt wurden – mit den entsprechenden Anpassungen hinsichtlich Synchronisationspunkt für die zeitliche Reorganisation der Daten und der Auswahl der IAs.

Die Auswertung von Daten aus Eyetracking-Experimenten im Visual-World-Paradigma ist derzeit allgemein ein großes Thema in der kognitionswissenschaftlichen bzw. psycholinguistischen Forschungsgemeinde. Zentrales Problem ist, dass in diesem Experimentalparadigma andere Datenrelationen als in üblichen psychologischen Experimenten mit kategorialer unabhängiger Prädiktorvariable und kontinuierlicher abhängiger Variable (z.B. Reaktionszeiten) herrschen. Wir haben zwar kategoriale Prädiktor-Variablen (unsere experimentellen Faktoren Anapherntyp und Satztyp), diese haben aber einen kontinuierlichen Verlauf über die Zeit (Fixationen über den Verlauf des Satzes). Des Weiteren sind unsere Ergebnisdaten nicht kontinuierlicher, sondern kategorialer Natur: Die abhängige Variable

ist der Anteil an Fixationen in ihrer Verteilung über eine begrenzte Anzahl von IAs – in unserem Experiment die Fixationswahrscheinlichkeit für die drei abgebildeten Personen.

In diesem Zusammenhang wurden verschiedene Verfahren vorgeschlagen. Allerdings herrscht nach wie vor Uneinigkeit über die Auswahl und Vorbereitung der Daten und über das Verfahren ihrer inferenzstatistischen Analyse. Als geeignete Verfahren wurden zum Beispiel Multilevel Logistic Regression (Barr, 2008), Growth Curve-Analysis (Mirman, Dixon & Magnuson, 2008) und Logit Mixed Models (T. F. Jaeger, 2008) thematisiert. Als Hauptproblem wird dabei die Analyse kategorialer Daten behandelt, für die die Varianzanalyse (ANOVA) ungeeignet ist. Auch die geeignete Erfassung des zeitlichen Verlaufs ist Thema der Debatte. In einer vergleichsweise pragmatischen Analyse berechnen Kaiser et al. (2009) die relativen Anteile der Fixationen auf den IAs für sukzessive Zeitscheiben der Größe 400 ms ab einem festgelegten On-Set (200 ms nach On-Set der anaphorischen Form). Für diese Zeitscheiben wurden jeweils unabhängig eine ANOVA durchgeführt. Die Verwendung logistischer Regressionsmodelle wird aufgrund ungeklärter Fragen betreffend ihrer Anwendung auf Visual-World-Daten abgelehnt. Eine nähere Erklärung dieses Sachverhalts bleibt leider aus, die Autoren verweisen auf Tanenhaus, Frank, Salverda, Jaeger und Masharov (2008).

### 7.3.4.1. Fixationsanteile auf den IAs

In einer ersten Analyse vergleichen wir die Fixationsanteile auf den IAs getrennt für die Bedingung des Anaphertyps.<sup>35</sup> Die Sekundär-Aufgabe ging nicht in die Analysen ein.<sup>36</sup> Die Fixationen wurden dazu Zeitscheiben der Größe 100 ms zugewiesen, anschließend wurde innerhalb der Zeitscheiben der relative Anteil der Fixationen für die drei IAs berechnet. Die Synchronisation des zeitlichen Verlaufs über die Items erfolgt am On-Set des anaphorischen Ausdrucks.<sup>37</sup> Die Zeitscheiben wurden ab 2500 ms vor und bis 4000 ms nach dem Synchronisationspunkt zugewiesen.

Wir haben die Fixationsanteile auf den drei IAs in zwei Schaubildern abgetragen, eines für Reflexivpronomen und eines für Personalpronomen. Die Linien tragen das arithmetische

---

<sup>35</sup>Die automatische Aufbereitung, Auswertung und graphische Darstellung erfolgte mit dem in R erstellten Tool *fix2bins.R*. Dieses wurde von Lars Konieczny für Visual-World-Experimente im Rahmen des ELoC-Projekts erstellt und für das Projekt AnaLoCo zur Verfügung gestellt.

<sup>36</sup>Zur Erinnerung: Sekundär-Aufgabe war es, diejenige Person anzuklicken, die ihrer Meinung nach in den beschriebenen Ereignissen die wichtigste Rolle spielt. In Bezug auf die Ergebnisse dieser Aufgabe hatten wir keine Hypothesen, da (entsprechend der Ergebnisse aus Experiment 1 für die Beantwortungsaufgabe) für die angenommenen Effekte lokaler Kohärenz nur eine zeitlich begrenzte Auswirkung auf die Interpretation erwartet wurde. Zudem war die Aufgabe so gestellt, dass es keine im eigentlichen Sinne ‚richtige‘ Antwort gab und somit aus den Ergebnissen auch keine Schlussfolgerungen gezogen werden können.

<sup>37</sup>Eine Synchronisation wird hier deshalb benötigt, da sowohl über die verschiedenen Sätze als auch innerhalb eines Satzes über die verschiedenen Bedingungen hinweg das Auftreten der für uns relevanten Satzposition an verschiedenen Stellen erfolgt. Um den Effekt der Interpretation für die Anapher beurteilen zu können, benötigen wir einen gemeinsamen Ausgangspunkt, über den wir kontrollieren können, dass wir die gleiche Satzposition vergleichen. Dies wird hier durch eine Verschiebung der Anfangszeit jeder einzelnen Fixation relativ zum Synchronisationspunkt angepasst, alle Fixationen werden also so verschoben, dass der phonetische On-Set der Anaphern übereinanderliegt. Da die Sätze aber auch nach dem anaphorischen Ausdruck wieder unsynchron werden, können wir nur Effekte beurteilen, die recht zeitnah zum Synchronisationspunkt auftreten. Folglich haben wir keine Datenpunkte in der Analyse berücksichtigt, die weiter als 2000 ms vom Synchronisationszeitpunkt entfernt sind.

Mittel des Fixationsanteils auf je einer Interest Area in Abhängigkeit zur Satzbedingung (mit oder ohne LSK) ab. Das transparente Feld um den Mittelwert repräsentiert den Standardfehler. Die vertikale gestrichelte Linie auf der x-Achse (Zeitachsenabschnitt 0) markiert den Synchronisationspunkt. Der Linientyp unterscheidet die Satzbedingung, die den Fixationen zugrundeliegt:

- Die durchgängige Linie (mit gefüllten Mittelwert-Markierungen der Zeitscheiben) entspricht dabei jeweils der Bedingung mit LSK (...*der Sohn kämmt sich...* und ...*der Sohn kämmt ihn...*)
- Die gestrichelte Linie (mit leeren Mittelwert-Markierungen der Zeitscheiben) der Kontrollbedingung (...*der Sohn gründlich kämmt sich...* und ...*der Sohn gründlich kämmt ihn...*)

Die Linienfarben (sowie die Markierungen der Zeitscheiben) unterscheiden die IAs:

- Die schwarze Linie (Mittelwert-Markierungen der Zeitscheiben: Quadrate) repräsentiert die Fixationsanteile auf dem Subjekt des Matrixsatzes (im Beispiel der VATER, Abb. 7.6 Mitte oben)
- Die dunkelgraue Linie (Mittelwert-Markierungen der Zeitscheiben: Rauten) repräsentiert die Fixationsanteile auf dem Subjekt des Relativsatzes (im Beispiel der SOHN, Abb. 7.6 rechts unten)
- Die hellgraue Linie (Mittelwert-Markierungen der Zeitscheiben: Dreiecke) repräsentiert den nicht benannten dritten Referenten (im Beispiel der CLOWN, Abb. 7.6 links unten)

Abbildung 7.7 zeigt die Fixationsanteile für die Bedingungen mit Reflexivpronomen. Wie für diese Bedingung vorhergesagt, finden wir eine signifikant höhere Fixationswahrscheinlichkeit auf dem Referenten des Relativsatz-Subjekts in der LSK-Bedingung als für die Kontrollbedingung. Der Effekt setzt etwa 600 ms nach On-Set des Reflexivpronomens ein und endet nach ungefähr 1600 ms. Berücksichtigt man dabei noch die Äußerungsdauer des Reflexivpronomens, erhalten wir einen Effektbeginn etwa 400 ms nach Off-Set des Pronomens. Unsere Unterschiedsbeurteilung erfolgt hier anhand des Unterschieds der beiden Mittelwerte: Die Fehlerbalken für die Standardfehler der beiden Mittelwerte überschneiden sich nicht, es zeigt sich zudem ein deutlicher Abstand zwischen ihren Extremen. Dies indiziert einen signifikanten Unterschied der beiden Bedingungen. Es zeigen sich auch Effekte in den anderen IAs. Zwar sind diese nur kurzlebig und von geringerem Unterschied, so dass wir diese Effekte vorerst nicht interpretieren können, allerdings unterstützt die Tendenz der Effektrichtung auf diesen IAs unsere Hypothesen.

Abbildung 7.8 zeigt die Fixationsanteile für die Bedingungen mit Personalpronomen. Wir finden hier weder auf dem Referenten des Relativsatz-Subjekts noch auf dem Referenten des Matrix-Subjekts einen signifikanten Unterschied. In dieser Hinsicht wurden unsere Hypothesen also nicht bestätigt. Es findet sich ein Unterschied für die Interest Area des unbeteiligten Referenten, der allerdings wiederum sehr kurzlebig ist (ab ca. 800 ms bis 1200 ms nach On-Set des Pronomens). Dieser Effekt einer Steigerung der Fixationsanteile entspricht unseren Hypothesen. Allerdings wird nicht ersichtlich, auf wessen Kosten diese Verschiebung der Fixationsanteile erscheint.

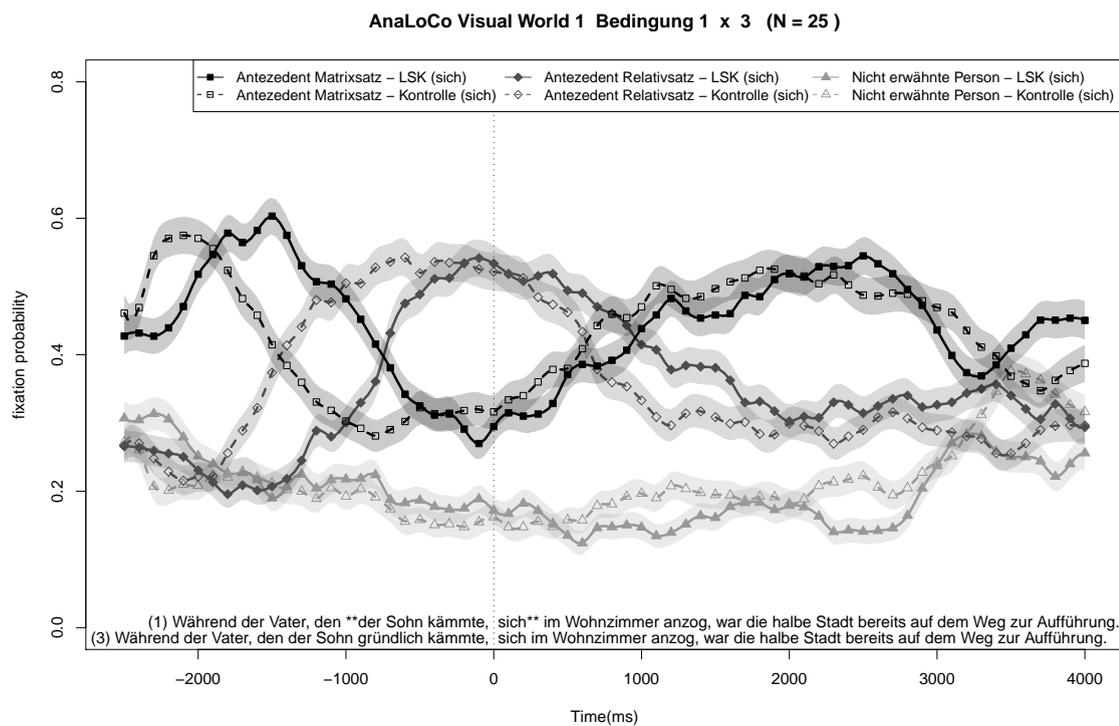


Abbildung 7.7.: AnaLoCo 2. Zeitlicher Verlauf der Fixationsanteile auf den drei IAs für die Bedingungen LSK vs. Kontrolle in den Versionen mit Reflexivpronomen. Relevant sind die Fixationsanteile nach dem Synchronisationszeitpunkt (vertikale gestrichelte Linie), ab dem übereinstimmende Stimuli-Positionen betrachtet werden.

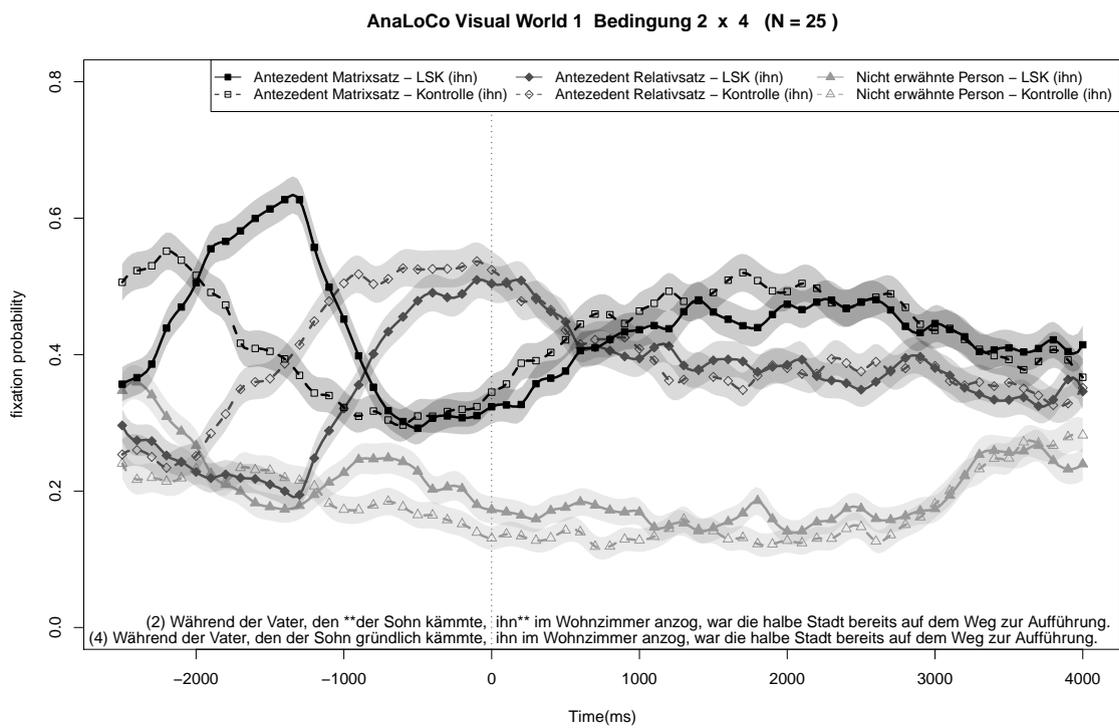


Abbildung 7.8.: AnaLoCo 2. Zeitlicher Verlauf der Fixationsanteile auf den drei IAs für die Bedingungen LSK vs. Kontrolle in den Versionen mit Personalpronomen. Relevant sind die Fixationsanteile nach dem Synchronisationszeitpunkt.

### 7.3.4.2. Gemischte lineare Modelle

Wir haben zur weiteren Überprüfung dieser Ergebnisse mit der Funktion *lmer* gemischte lineare Modelle für die Daten gefittet. Der von uns gewählte Ansatz für die Inferenzstatistik ist mit dem Vorgehen von Kaiser et al. (2009) vergleichbar. Grundlage der Analyse sind die relativen Anteile der Gesamtfixation pro Zeitscheibe für die gewählten IAs, allerdings verwenden Kaiser und Kollegen ANOVAs anstelle von linearen gemischten Modellen. Wir haben getrennte Modelle für die beiden Anaphertypen berechnet. In die Modelle ging jeweils als fester Faktor (fixed effects) der kontrollierte Experimental-Faktor Satztyp hinsichtlich seiner Interaktion mit den drei IAs ein. Wir erwarten also keinen Haupteffekt für die Bedingung des Satztyps, und ein Haupteffekt für die IAs ist trivial und ohne Aussage. Vielmehr geht es darum, wie der Faktor Satztyp den Anteil der Fixationen auf den IAs beeinflusst. Als Zufallsfaktoren (random effects) haben wir die Intercepts für VTn und Items aufgenommen.

Da die Fixationen einem zeitlichen Verlauf folgen, mussten wir uns für ein Verfahren zur Abgrenzung des zu untersuchenden Abschnitts entscheiden. Dabei versuchen wir, der Analyse möglichst große Zeitscheiben (600-1000 ms) zugrunde zu legen, um zu verhindern, dass über eine willkürliche Auswahl des Zeitabschnitts zufällige Scheineffekte (im Sinne eines  $\alpha$ -Fehlers) berücksichtigt werden. Wir verwenden für die folgende Analyse der Bedingung mit Reflexivpronomen das arithmetische Mittel der Fixationswahrscheinlichkeit für neu berechnete Zeitscheiben mit einer Länge von 1000 ms. In der Bedingung mit Personalpronomen haben wir eine kleinere Zeitscheibe von 600 ms verwendet. In beiden Fällen haben wir für den On-Set der Zeitscheiben den Zeitabschnitt ab 500 ms nach dem On-Set der Anapher gewählt.

Die beiden Diagramme 7.9 und 7.10 zeigen das arithmetische Mittel und den Standardfehler für die gewählten Zeitscheiben in den beiden Bedingungen. Die darauffolgenden Tabellen 7.8 und 7.9 stellt den Term des jeweils dafür berechneten Modells, die Ergebnisse für feste Faktoren und Zufallsfaktoren, sowie die Schätzung für F- und p-Werte für Haupteffekte und Interaktionen der festen Faktoren über die Funktion *aovlmer.fnc* dar.<sup>38</sup>

Für die Reflexivpronomen findet sich eine signifikante Interaktion zwischen Interest Area und Satzbedingung ( $F = 4.39$ ,  $p < 0.05$ ). Diese ist auf den Effekt erhöhter Fixationsanteile auf der Abbildung des Relativsatz-Subjekts in der LSK-Bedingung zurückzuführen ( $t = 2.159$ ). Zu den Haupteffekten können in diesen Modellen keine sinnvollen Aussagen über Unterschiede für die Satzbedingungen getroffen werden.<sup>39</sup> Der höchst signifikante

---

<sup>38</sup>Eine kurze Erläuterung der Variablen in den folgenden Tabellen: mean\_POI = Fixationsanteile; Condition = Bedingungen des Faktors Satztyp; Condition-Kontrolle = Satzbedingung ohne LSK (Ausgangspunkt des statistischen Vergleichs); Condition-LSK = Bedingung mit LSK; IA = Interest Area; IA-Antezedens(MC) = IA für das Matrixsatz-Subjekt (Ausgangspunkt des statistischen Vergleichs); IA-Antezedens(RC) = IA für das Relativsatz-Subjekt; IA-Non-Antezedens = IA für die nicht benannte Person; Item\_ID = Identität des Stimulussatzes; VTn = Identität des Versuchsteilnehmers.

<sup>39</sup>Der Haupteffekt für IAs ist für die Analyse der Unterschiede irrelevant, da er trivialerweise nur aussagt, dass es auf der Zeitscheibe unterschiedliche Fixationsanteile für die Abbildungen gibt. Er ist aber insofern interessant, als er anzeigt, auf welcher Abbildung sich absolut die höchsten Fixationsanteile finden. Für die Bedingungen kann es keinen Effekt geben, da die Fixationsanteile nicht unabhängig voneinander sind und sich die Verhältnisse somit ausgleichen: Erhält eine visuelle Region innerhalb einer Bedingung mehr Fixationen, so müssen sich zwingend auf einer anderen Region in derselben Bedingung weniger Fixationen finden.

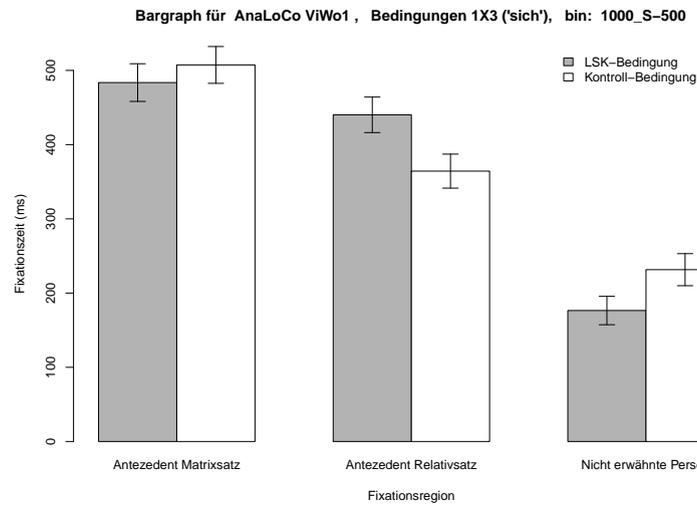


Abbildung 7.9.: AnaLoCo 2. Balkendiagramm für die Bedingung Reflexivpronomen: Mittelwert der Fixationsanteile (y-Achse) für eine 1000 ms-Zeitscheibe (ab 500 ms nach On-Set der Anapher) auf den drei IAs (x-Achse) in Abhängigkeit der Satzbedingung. Fehlerbalken repräsentieren den Standardfehler.

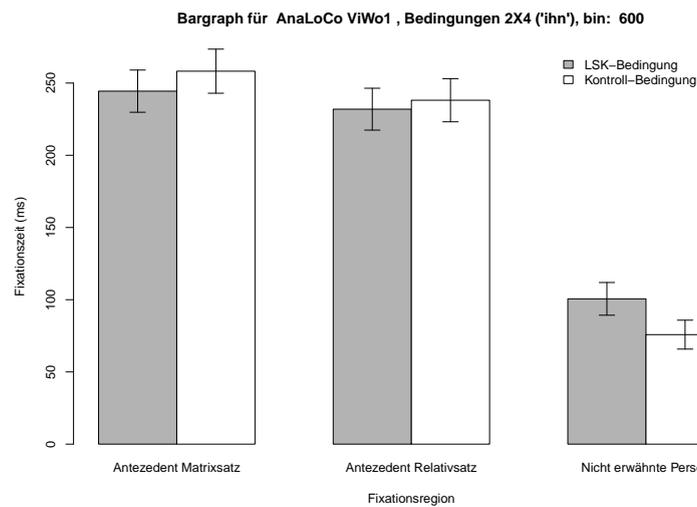


Abbildung 7.10.: AnaLoCo 2. Balkendiagramm für Bedingung Personalpronomen: Mittelwert der Fixationsanteile (y-Achse) für eine 600 ms-Zeitscheibe (ab 500 ms nach On-Set der Anapher) auf den drei IAs (x-Achse) in Abhängigkeit der Satzbedingung. Fehlerbalken repräsentieren den Standardfehler.

Linear mixed model fit by REML

Formula: mean\_POI ~Condition \* IA + (1 | Item\_ID) + (1 | VTn)

Random effects:

Groups	Name	Variance	Std.Dev.
Item_ID	(Intercept)	0	0.0
VTn	(Intercept)	0	0.0
Residual		159042	398.8

Number of obs: 1794, groups: Item\_ID, 48; vp, 25

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	507.36	23.06	21.999
Condition-LSK	-23.77	32.62	-0.729
IA-Antezedens(RC)	-143.07	32.62	-4.386
IA-Non-Antezedens	-275.84	32.62	-8.457
Condition-LSK:IA-Antezedens(RC)	99.60	46.13	2.159
Condition-LSK:IA-Non-Antezedens	-31.26	46.13	-0.678

Analysis of Variance Table:

	Df	Sum Sq	Mean Sq	F value	F	Df2	p
Condition	1	441	441	0.0028	0.0028	1788	0.95803
IA	2	26497867	13248934	83.3046	83.3046	1788	0.00000
Condition:IA	2	1396420	698210	4.3901	4.3901	1788	0.01253

Tabelle 7.8.: AnaLoCo 2. Gemischtes lineares Modell für den Faktor Satztyp in der Bedingung Reflexivpronomen, Fixationsanteile auf den IAs für eine Zeitscheibe von 1000 ms ab einem On-Set 500 ms nach dem Synchronisationspunkt.

Haupteffekt für den Faktor Interest Area ( $F = 83.3046$ ,  $p < 0.0001$ ) weist darauf hin, dass die Abbildungen mit deutlich unterschiedlicher Wahrscheinlichkeit betrachtet werden. Die t-Werte zeigen, dass das Agens des Matrixsatzes den größten Anteil beansprucht und das Agens des Relativsatzes einen signifikant größeren Anteil erhält als der unbeteiligte Referent.

Der erwartete Effekt für die Version mit Personalpronomen (erhöhte Fixationsanteile auf der Abbildung des unbeteiligten Agens) ließ sich im Modell (Tab. 7.9) zunächst nicht bestätigen (Interaktion von Satztyp und Anapherntyp:  $F = 1.1315$ ,  $p > 0.32$ , t-Wert für den Vergleich der Interaktion zwischen dem unbeteiligten Referenten und der globalen Bedingung:  $t = 1.42$ ). Auch für kleinere Zeitscheiben und variable On-Sets wurde kein Effekt gefunden. Allerdings messen wir mit den verwendeten Modellen nur Interaktionen, nicht aber Haupteffekte für die einzelnen IAs. Wenn wir jedoch ein Modell für die Satztyp-Bedingungen isoliert auf der Interest Area des unbeteiligten Referenten berechnen (vgl. Tab. 7.10), finden wir den Effekt in erwarteter Wirkungsrichtung ( $t = 1.966$ ,  $F = 3.8652$ ,  $p < 0.5$ ). Der höchst signifikante Haupteffekt für den Faktor Interest Area ( $F = 87.3984$ ,  $p < 0.0001$ ) beruht den t-Werten folgend auf einem deutlich geringeren Anteil des unbeteiligten Agens an den Fixationen. Dieser Effekt war zu erwarten.

### 7.3. AnaLoCo 2: Visual-World-Experiment mit Referenten

Linear mixed model fit by REML

Formula: mean\_POI ~Condition \* IA + (1 | Item\_ID) + (1 | VTn)

Random effects:

Groups	Name	Variance	Std.Dev.
Item_ID	(Intercept)	0	0.00
VTn	(Intercept)	0	0.00
Residual		55034	234.59

Number of obs: 1788, groups: Item\_ID, 48; vp, 25

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	258.221	13.590	19.001
Condition-LSK	-13.832	19.219	-0.720
IA-Antezedens(RC)	-20.121	19.219	-1.047
IA-Non-Antezedens	-182.409	19.219	-9.491
Condition-LSK:IA-Antezedens(RC)	7.648	27.179	0.281
Condition-LSK:IA-Non-Antezedens	38.607	27.179	1.420

Analysis of Variance Table:

	Df	Sum Sq	Mean Sq	F value	F	Df2	p
Condition	1	1125	1125	0.0204	0.0204	1782	0.88635
IA	2	9619810	4809905	87.3984	87.3984	1782	0.00000
Condition:IA	2	124540	62270	1.1315	1.1315	1782	0.32279

Tabelle 7.9.: AnaLoCo 2. Gemischtes lineares Modell für den Faktor Satztyp in der Bedingung Personalpronomen, Fixationsanteile auf den IAs für eine Zeitscheibe von 600 ms ab einem On-Set 500 ms nach dem Synchronisationspunkt.

Linear mixed model fit by REML

Formula: mean\_POI ~Condition + (1 | Item\_ID) + (1 | VTn)

Random effects:

Groups	Name	Variance	Std.Dev.
VTn	(Intercept)	12917.2	113.654
Item_ID	(Intercept)	2599.8	50.988
Residual		104871.0	323.838

Number of obs: 432, groups: vp, 36; Item\_ID, 24

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	351.68	30.88	11.387
Condition-LSK	61.26	31.16	1.966

Analysis of Variance Table:

	Df	Sum Sq	Mean Sq	F value	F	Df2	p
Condition	1	405353	405353	3.8652	3.8652	430	0.049939

Tabelle 7.10.: AnaLoCo 2. Gemischtes lineares Modell für den Faktor Satztyp in der Bedingung Personalpronomen, Fixationsanteile auf der Interest Area mit Abbildung des unbeteiligten Referenten für eine Zeitscheibe von 600 ms ab einem On-Set 500 ms nach dem Synchronisationspunkt.

### 7.3.5. Diskussion AnaLoCo 2

**Reflexivpronomen** Die Ergebnisse des ersten Visual-World-Experiments sind vielversprechend. Im Fall des Reflexivpronomens geben sie einen Hinweis darauf, dass es sich bei den in Experiment 1 gefundenen Effekte nicht nur um Interferenzen der lokalen mit der globalen Lesart auf syntaktischer Ebene handelt. Vielmehr deuten sie auf eine durch LSK induzierte temporäre Verfügbarkeit des strukturell unzugänglichen lokalen Antezedens hin. Dabei bestätigen sich unsere Hypothesen in Bezug auf die erhöhten Fixationsanteile auf dem Subjekt des Relativsatzes. Dies legt unsere Vermutung nahe, dass unter Verletzung der globalen Kohärenz bzw. unter Verletzung der globalen Bindungsdomänen innerhalb der LSK das lokale Agens als Antezedens des Reflexivpronomens in Betracht gezogen wird. Allerdings lässt sich aus den Daten nicht bestätigen, dass diese Fixationen vom Subjekt des Matrixsatzes abgezogen werden – eher weist die Abbildung auf eine leichte Tendenz eines Abzugs von der Abbildung des unbeteiligten Referenten hin. Der Effekt findet sich wie erwartet sehr früh (ab ca. 500 ms) und ist persistent bis etwa 1700 ms nach On-Set der Anapher.

Eine mögliche Alternativerklärung des LSK-Effekts für Reflexivpronomen, die ohne die Annahme der tatsächlichen Interpretation der LSK auskommt, würde die Daten so deuten, dass initial auch für diese Klasse von Anaphern immer auch strukturell unpassende Referenten berücksichtigt werden. Auf eine solche initial verminderte Einschränkung weisen Runner et al. (2006) hin. Aufgrund der strukturell ambigen Situation in der LSK-Bedingung würde man dabei erwarten, dass es länger dauert, bis der strukturelle Filter die globale Interpretation durchsetzen kann. Dies würde bedeuten, dass sich die LSK nicht bis auf die Ebene der Referenzzuweisung auswirken kann, sondern nur ein präinterpretativer Interferenz-Effekt ist.

**Personalpronomen** Auch für die Personalpronomen wurden Hinweise auf einen Effekt einer vermehrten Fixation des unbeteiligten Referenten gefunden. Diese vermehrte Aufmerksamkeit deutet an, dass temporär eine Suche nach einem alternativen Antezedens einsetzt, da das bereits genannte lokale Antezedens innerhalb der LSK fälschlicherweise strukturell ausgeschlossen wird. Unklar ist allerdings, warum für den Effekt keine Interaktion gefunden wurde, denn die Tendenzen der anderen beiden kontrastierenden IAs weisen schließlich in eine entgegengesetzte Effektrichtung. Offen bleibt auch die Frage, von wem die visuelle Aufmerksamkeit abgezogen wird. Des Weiteren konnte der Effekt nicht für eine größere Zeitscheibe bestätigt werden. Er findet sich nur auf der begrenzten Zeitscheibe von 600 ms. Hinzu kommt, dass wir keinen Effekt temporär reduzierter Aufmerksamkeit auf dem lokalen Antezedens finden konnten. Dieser sollte innerhalb der LSK strukturell als Antezedens des Personalpronomens ausgeschlossen sein. Möglicherweise wird der Kreis potenzieller Antezedenzen bei ambigem strukturellem Kontext ausgeweitet, und für die Auflösung der anaphorischen Referenz werden schon frühzeitig andere (z.B. morphosyntaktische, semantische oder diskursbezogene) Constraints herangezogen.

Wie der Effekt für die Version mit Reflexivpronomen zeigt, scheinen die Stimuli und das verwendete Verfahren sensitiv genug zu sein, um LSK-Effekte zu messen. Wieso sind aber die Befunde für die Personalpronomen so uneindeutig? Eine Möglichkeit wäre, dass sich unsere strukturelle Manipulation auf die Effekte für diese Klasse von Anaphern

nicht so geradlinig bzw. vorhersagbar auswirkt, wie dies bei den Reflexivpronomen der Fall ist. Dies würde sich mit Befunden von Kaiser et al. (2009) und Runner et al. (2006, 2006) decken, die eine unterschiedliche Sensitivität verschiedener anaphorischer Klassen für unterschiedliche Faktoren beschreiben. Während Reflexivpronomen sehr empfänglich für strukturelle Faktoren sind, wird die Bindung von Personalpronomen sehr viel stärker durch semantische und pragmatische Faktoren bedingt. Demnach sollten sich strukturelle Interferenzen stärker auf Reflexiv- als auf Personalpronomen auswirken. Da unsere Stimuli keine weitergehenden semantischen Hinweise auf das korrekte Antezedens zur Verfügung stellen und die strukturellen Hinweisreize temporär ambig erscheinen, finden wir auf gegebener Satzposition nur eine unterspezifizierte Auswahl aller gegenwärtigen morphosyntaktisch kongruenten Referenten. Folglich finden wir eine stärkere Verteilung der visuellen Aufmerksamkeit auf diesen.

Eine ganz andere Möglichkeit hängt mit der Frage zusammen, durch welche Aspekte der sprachlichen Stimuli die visuelle Aufmerksamkeit in welcher Weise geleitet wird. So können wir beispielsweise nicht mit Sicherheit davon ausgehen, dass bei der Nennung der Anapher ausgerechnet das Antezedens derselben fixiert wird – und damit das aktuell benannte Objekt des Satzes. Gleichmaßen könnte das Agens der beschriebenen Handlung im Fokus der Aufmerksamkeit stehen. Damit würde anstelle des aktuellen Objekts die aktuelle Handlung des zuletzt benannten Verbs die Aufmerksamkeit steuern. Im Fall des Reflexivpronomens fallen Agens und Antezedens zusammen. In beiden Fällen können wir hier einen Effekt der LSK erwarten. Im Fall des Personalpronomens ist das Agens der transitiven Handlung der *Sohn* – sowohl in der LSK- als auch in der Kontrollbedingung. Denn dieser ist das Subjekt des Objekt-Relativsatzes, der die transitive Handlung beschreibt. Somit unterscheidet sich zwar temporär das Antezedens des Personalpronomens in den beiden Bedingungen, das Agens der Handlung bleibt aber dasselbe. Folglich können wir in diesem Fall keinen Unterschied in den Fixationsanteile finden. Eine andere Alternative mit demselben Ergebnis wäre, dass sich die visuelle Aufmerksamkeit in transitiven Situationen gleichermaßen (oszillierend) sowohl auf das Agens als auch auf das Patiens richtet. In diesem Fall sollten sich die Fixationen ausgleichen.

**Probleme und Einschränkungen** Da die Befunde für die Reflexiv- und auch für die Personalpronomen darauf hinweisen, dass abhängig von der syntaktischen Struktur potenzielle Referenten bzw. Kandidaten für die Argumentstruktur mit unterschiedlicher Wahrscheinlichkeit in Betracht gezogen werden, findet die Annahme weitere Bestätigung, dass LSK tatsächlich Einfluss auf die referenzielle Interpretation nehmen. In Abhängigkeit davon, was die Wahl des Aufmerksamkeitsfokus determiniert, besteht die Möglichkeit, dass die Fixationsanteile über einzelne Referenten ohne Handlungskontext und damit ohne zugewiesene thematische Rollen für Personalpronomen nicht so informativ sind wie für Reflexivpronomen. Dies ist das erste fundamentale Problem, das in unserem Folgeexperiment (ebenfalls im Visual-World-Paradigma) durch eine alternative Wahl der visuellen Stimuli thematisiert wird.

Ein zweiter Sachverhalt, der ebenfalls einer besseren Kontrolle bedarf, betrifft die Wahl der sprachlichen Stimuli. Die Durchbrechung der LSK-Lesart in den Kontrollsätzen wurde in den Experimenten 1 und 2 dadurch konstruiert, dass in den Relativsatz ein Adverb eingefügt wurde. Dies bringt zwei Probleme mit sich: Erstens erhalten die kontrastierten

Sätze von der kritischen Position der Anapher eine unterschiedliche Länge und beinhalten unterschiedliches Material, das die Lesezeiten zusätzlich verändern könnte. Zwar werden die Längenunterschiede durch die Synchronisierung der Daten anhand des On-Sets des Verbs kontrolliert. Eine methodisch sauberere Gestaltung der Kontrollstimuli wäre es aber, wenn wir für diese Sätze eine zumindest annähernd gleiche Länge wie bei den jeweiligen Zielsätzen erreichen. Zweitens führt das Adverb eine alternative lokale Kohärenz ein, die zwar auf einer seltenen Satzstruktur mit topikalisiertem Adverb (*Gründlich kämmt ihn...*) beruht, aber dennoch einen eigenen Einfluss auf die Interpretation des Satzes und somit auf die Fixationsanteile nehmen könnte.

Bezüglich des Problems der Längenunterschiede muss auch noch eine Alternativerklärung der Effekte erwähnt werden. Betrachtet man den Entwicklungsverlauf der Fixationsanteile, könnte man folgendes unterstellen: Der vermeintliche Effekt der Satztyp-Bedingungen nach dem Synchronisationspunkt entsteht durch die Verschiebung der Fixationsverläufe, die durch die unterschiedlichen Satzlängen vor dem Synchronisationspunkt entsteht (zumindest im Fall der Reflexivpronomen, vgl. Abb. 7.7). Hierfür wäre folgende Erklärung denkbar: Für die Fixation einer möglichen Szene ist eine Mindestzeit erforderlich, um die Korrespondenz zu bestätigen. Das bedeutet, nachdem der Blick auf eine Abbildung gerichtet wurde, verweilt er für einen Mindestzeitraum darauf. Da die Zeit bis zur erforderlichen Reorientierung der visuellen Aufmerksamkeit am Übergang zum Matrixsatz im Fall der LSK-Bedingung geringer ist, verweilt die Aufmerksamkeit nach dem Synchronisationspunkt noch etwas länger auf dem zuvor betrachteten Bild, bis eine Mindestfixationszeit erreicht wurde. Entgegen dieser Alternativerklärung gehen wir aber davon aus, dass die verfügbare Zeit in beiden Fällen genügt, und dass die Effekte durch die Synchronisation der Fixationen tatsächlich den intendierten experimentellen Manipulationen zuzuschreiben sind. Um aber ausschließen zu können, dass die Effekte durch einen bloßen Satzlängenunterschied zustande kommen, wird es umso wichtiger, die Kontrolle der sprachlichen Stimuli zu verbessern.

Ein drittes Problem neben der Kritik der Abbildungen und der sprachlichen Stimuli betrifft die Konstruktion der auditiven Stimuli aus dem Satzmaterial – das Einschneiden der Zielstruktur, mit dem eine Manipulation bzw. Zerstörung der prosodischen Markierung einhergeht. Dieses Problem werden wir in dieser Studie nicht beseitigen. Im folgenden Unterkapitel werden wir kurz diskutieren, weshalb wir uns für diese Experiment-Gestaltung entschieden haben.

Trotz dieser Einschränkungen und der Unklarheit, was die Experimentalbedingungen genau bewirken, können wir festhalten, dass die LSK die Lenkung der visuellen Aufmerksamkeit beeinflussen. Unter der Annahme, dass die Unterschiede in der visuellen Aufmerksamkeit auf den Unterschieden der Interpretationen des gehörten Satzes beruhen, können wir daraus schließen, dass die Verarbeitung von LSK bis zur Ebene der Interpretation durchschlägt.

### **7.3.6. Auswirkung prosodischer Manipulationen**

Eine grundlegende Schwäche des hier in den Visual-World-Experimenten verfolgten Ansatzes, die wir aber von vornherein im Design einkalkuliert haben, liegt in der Gestaltung

der Stimuli. Wir verwenden geschnittene Audio-Materialien. Dabei werden die LSK in möglichst neutraler Form in den Satz eingeschnitten, wodurch prosodische Merkmale wie zum Beispiel die Grenze des Relativsatzes nivelliert werden. Dies ist ganz klar eine Manipulation, die den LSK-Effekt begünstigt. Denn es ist unwahrscheinlich, dass die Hauptsatz-Lesart in Betracht gezogen würde, wenn durch eine Anhebung, eine Absenkung oder sonstige Intonationsmarkierungen des Teilsatzendes sowie durch eine Pause nach der Teilsatzgrenze die lokale Satzstruktur durchbrochen wird. Im schlimmsten Fall werden die VTn mit den Materialien sogar zur Verarbeitung der LSK gedrängt, da diese eine prosodische Markierung erwarten – ihr Ausbleiben wird als Hinweis auf die Fortsetzung der aktuellen Satzstruktur interpretiert. Zu vermuten ist jedenfalls, dass prosodische Markierungen die LSK genauso zerstören wie nicht kohärente Wörter (etwa das Adverb in unserer Kontrollbedingung).

Die Entscheidung, auditive Materialien zu verwenden, die durch Einschneidungen manipuliert wurden, wurde bewusst getroffen. Die Studie gilt nicht der Untersuchung, wie syntaktische Analyse durch die Verarbeitung prosodischer Markierung unterstützt wird. Vielmehr geht es darum, Verarbeitungseigenschaften der syntaktisch-strukturellen Analyse – basierend auf Wörtern, Phrasen und morphosyntaktischen Markierungen – isoliert zu untersuchen. Denn genau an diesen Eigenschaften werden viele Annahmen der globalen Inkrementalität der Satzverarbeitung, der Kohärenz der aufgebauten Dependenzstruktur und eben der Bindung koreferierender Argumente festgemacht. Rein syntaktisch betrachtet sollte die lokal isolierte Lesart der LSK von vornherein ausgeschlossen sein – auch wenn prosodische Merkmale fehlen.

Dennoch kann kritisch diskutiert werden, inwiefern wir mit unseren Materialien natürliche Sprachverarbeitung untersuchen. Unbestritten hat Prosodie einen starken Einfluss auf die strukturelle Interpretation gehörter Sätze (vgl. Kap. 5.2.3). So zeigt etwa DeDe (2010), dass prosodische Merkmale zum frühestmöglichen Zeitpunkt Parsingpräferenzen beeinflussen und unmittelbar mit lexikalischen Informationen und Plausibilitäts-Constraints interagieren. Aus einer solchen Perspektive können LSK-Effekte möglicherweise als weiterer Beleg für die Bedeutung der Prosodie in der Satzverarbeitung verstanden werden und als Hinweis auf die unmittelbare parallele Integration multipler Constraints bei der Satzverarbeitung dienen. Konieczny, Müller, Baumann et al. (2009) untersuchen in einer Studie zu LSK den Einfluss der Prosodie mit zwei Experimenten, über die der Einfluss der Audio-Manipulation variiert wurde: (a) LSK in eingeschnittener LSK (manipulierte Prosodie) und (b) LSK in natürlicher Prosodie (nicht-manipulierte Variante). In der Variante mit natürlicher Prosodie wurde kein LSK-Effekt gefunden, während im auf gleichen Sätzen basierenden Experiment mit Audio-Manipulation Effekte für die LSK-Bedingungen gefunden wurden. Dies weist darauf hin, dass LSK-Effekte möglicherweise künstlicher Natur sind. Es zeigt aber auch die Bedeutung der Prosodie für eine reibungslose und fehlerfreie Satzverarbeitung sowie die Unmittelbarkeit der Integration dieser Merkmale.

#### 7.3.6.1. Normierungsstudie

Da durch die in Experiment 2 verwendete LSK eine Satzgrenze überschritten und eben diese Satzgrenze in den von uns konstruierten geschnittenen Audio-Materialien nivelliert wird, müssen wir vermuten, dass die verwendeten Materialien möglicherweise als unna-

türlich empfunden werden. Um zu überprüfen, inwiefern sich die geschnittenen auditiven Stimuli mit eingefügter LSK-Sequenz von unbehandelten Stimuli unterscheiden und ob diese Unterschiede tatsächlich wahrgenommen werden, haben wir eine Normierungsstudie durchgeführt. In dieser als Rating-Experiment durchgeführten Studie hören die Teilnehmer die Sätze aus verschiedenen Experimenten und geben nach jedem gehörten Satz eine Beurteilung über die Verständlichkeit und die prosodische Natürlichkeit des Audio-Stimulus ab. Ein Effekt für den Unterschied der Bearbeitung muss als mögliche Alternativerklärung der in Experiment 2 gefundenen Effekte in der Diskussion berücksichtigt werden.

**Design und Material** Wir verwenden in der Studie ein  $2 \times 4$ -Design. Der erste Faktor unterscheidet den Audiotyp in den Bedingungen *spliced* (eingeschnittene LSK) und *non-spliced* (nicht bearbeiteter Satz).<sup>40</sup> Der zweite Faktor unterscheidet die verwendeten Materialien, die wir vor der Besprechung des Faktors kurz skizzieren. Als Material haben wir die auditiven Stimuli aus verschiedenen Experimenten<sup>41</sup> zusammengestellt, für die sowohl geschnittene als auch ungeschnittene Versionen der Sätze bestanden. Aus dem gesammelten Satzmaterial wurde ein Subset ausgewählt (je 24 Sätze aus vier Experimenten), so dass insgesamt eine Liste aus 96 Satz-Items entstand.

In der Kodierung der Materialien wurden die vier zugrundeliegenden Experimente als Faktorstufen aufgenommen. Dabei haben wir das AnaLoCo-Experiment 2 in zwei Untergruppen für Reflexivpronomen und Personalpronomen geteilt. Die Aufnahme der Experimente als Faktor ermöglicht uns eine gleichmäßige Verteilung der Bedingungen über die Experimente in der Rotation (diese erfolgte über die Latin-Square-Methode), eine getrennte Untersuchung der Experimente und die Untersuchung von Rating-Unterschieden in Abhängigkeit des verwendeten Anaphertyps.

**Methode** Die Studie wurde mit der Experimental-Software Linger (Rohde, 1999a) implementiert. Zur Durchführung mussten sich die Teilnehmer die Sätze über Lautsprecher am PC anhören und durch anklicken von sogenannten Radio-Buttons (Optionsschaltflächen) beurteilen. Die Ratingskalen für die nachfolgende Beurteilung wurden auf dem Bildschirm eingeblendet. Aufgezeichnet wurde die Latenz der Beurteilung (ab Einblendung bis zur Beurteilung) sowie das Rating (der Beurteilungswert).

Insgesamt nahmen 31 VTn am Experiment teil, vorwiegend Studierende der Universität Freiburg. Das Experiment wurde mit anderen Experimenten zusammen in eine größere Studie eingebettet, über die die Vergütung der Teilnehmer erfolgte.

Die VTn mussten durch das Drücken einer Taste den ersten Satz starten und sich diesen anhören. Zentrale Aufgabe war es, zunächst den Inhalt des Satzes zu verstehen und dann

---

<sup>40</sup>Dem ist anzumerken, dass auch die nicht bearbeitete Variante tatsächlich ebenfalls leicht bearbeitet wurde. Allerdings wurde in der zweiten Variante eben nicht die LSK eingeschnitten, d.h. die natürliche Prosodie bleibt erhalten. In der Bedingung *non-spliced* wurde ausschließlich die Länge der Pausen zwischen Teilsätzen angleichend gekürzt. Durch diese Bereinigung sollten für unsere Fragestellung irrelevante Störfaktoren reduziert werden.

<sup>41</sup>Das Material entstammt zum einen der AnaLoCo-Studie Experiment 2 sowie den ungeschnittenen Versionen der darin verwendeten Sätze. Zum anderen stammt es aus früheren Experimenten des Projekts ELoC, die unter anderem in Konieczny, Müller, Hachmann et al. (2009) und Konieczny und Müller (2010) vorgestellt wurden.

die Verständlichkeit von Inhalt und Bedeutung des Satzes sowie die Natürlichkeit von Klang, Rhythmik und Betonung des Satzes zu bewerten. Nach der akustischen Präsentation des Satzes erschien erst die Befragung zur Verständlichkeit und danach die Befragung zur Natürlichkeit. Die Bewertung erfolgte auf einer Skala zwischen 1 (für sehr gut verständlich/sehr natürlich) bis 7 (sehr schlecht verständlich/sehr unnatürlich). Die Skalen wurden als horizontale Reihe von Radio-Buttons eingeblendet, die Ratings wurden durch Anklicken der gewählten Beurteilung mit der Maustaste vergeben. Direkt nach Beantwortung des zweiten Ratings (mit kurzer Verzögerung von 2 Sekunden) wurde automatisch der jeweils nächste Satz präsentiert.

### 7.3.6.2. Ergebnisse der Normierungsstudie

Die Ratings wurden z-standardisiert, die Latenzen der Beurteilung logarithmiert. Die folgenden Liniendiagramme geben die Ergebnisse der Ratingstudie wieder und zeigen die Unterschiede der Ratings für den Faktor Audiotyp sowie den Faktor Anapherntyp – ermöglicht durch die isolierte Betrachtung des gruppierten Experiments 2. Diagramm 7.11 zeigt die z-standardisierten Ratings für die Verständlichkeit, Diagramm 7.12 die z-standardisierten Ratings für die Natürlichkeit der Prosodie. Für die Latenzen bilden wir keine Diagramme ab.

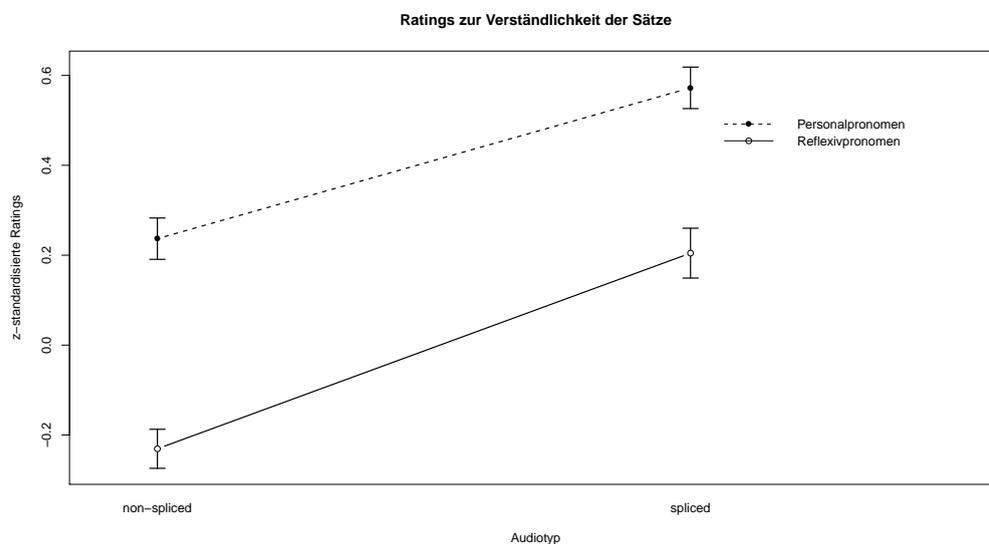


Abbildung 7.11.: Ergebnisse der Ratingstudie (z-standardisiert) für das Rating der Verständlichkeit der Sätze. Fehlerbalken repräsentieren den Standardfehler.

Wie aus Abbildung 7.11 ersichtlich wird, wird die Verständlichkeit sowohl von der Manipulation der Audio-Stimuli als auch vom Anapherntyp beeinflusst. So zeigen sich deutlich schlechtere Ratings für die geschnittenen Varianten (x-Achse) und auch für die Satzvarianten mit Personalpronomen (gestrichelte Linie, Mittelwerte dargestellt als gefüllte Kreise). Zur weiteren Prüfung der Ergebnisse wurde ein gemischtes lineares Modell

gefittet.<sup>42</sup> Das Modell für die z-standardisierten Ratings für die Verständlichkeit bestätigt den Befund aus dem Diagramm: Es indiziert einen höchst signifikanten Haupteffekt für den Faktor Audiotyp ( $t = 5.106$ ,  $F = 67.7251$ ,  $p < 0.0001$ ) und den Faktor Anapherentyp ( $t = -7.147$ ,  $F = 83.5501$ ,  $p < 0.0001$ ), aber keine Interaktion der beiden Faktoren ( $t = 1.023$ ,  $F = 1.0471$ ,  $p > 0.3$ ).

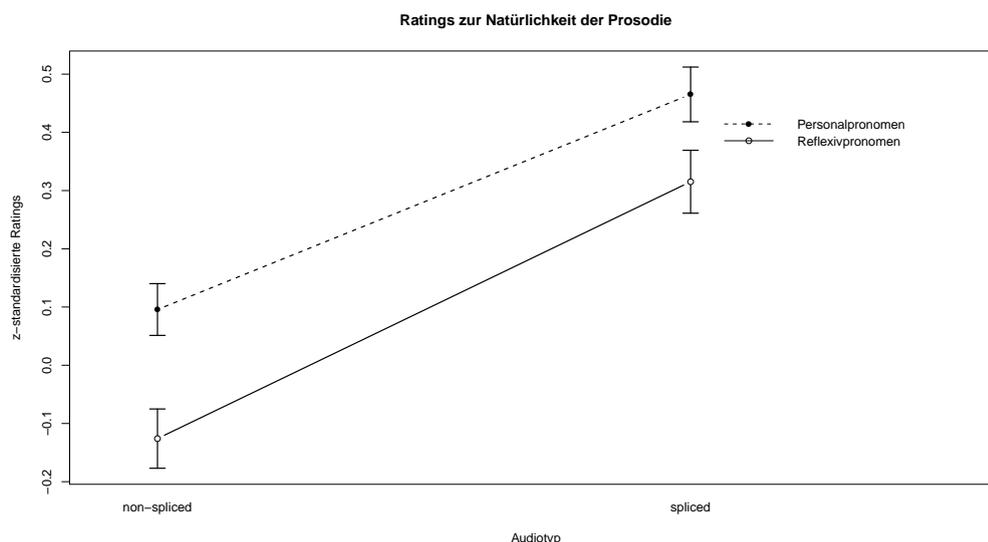


Abbildung 7.12.: Ergebnisse der Ratingstudie (z-standardisiert) für das Rating der Natürlichkeit der Prosodie. Fehlerbalken repräsentieren den Standardfehler.

Ein identisches Bild findet sich für die Beurteilung der Natürlichkeit der Prosodie: Hier ergeben sich deutlich schlechtere Ratings für die geschnittenen Varianten (x-Achse) und auch für die Satzvarianten mit Personalpronomen (gestrichelte Linie, Mittelwerte dargestellt als gefüllte Kreise). Zur weiteren Prüfung der Ergebnisse wurde auch hierfür ein gemischtes lineares Modell gefittet.<sup>43</sup> Wieder bestätigt das Modell für die z-standardisierten Ratings für die Natürlichkeit der Prosodie den Befund aus dem Diagramm. Es zeigt einen höchst signifikanten Haupteffekt für den Faktor Audiotyp ( $t = 5.454$ ,  $F = 73.1041$ ,  $p < 0.0001$ ) und den Faktor Anapherentyp ( $t = -3.414$ ,  $F = 15.8687$ ,  $p < 0.0001$ ), aber wiederum keine Interaktion der beiden Faktoren ( $t = 0.846$ ,  $F = 0.7161$ ,  $p > 0.39$ ).

Die Übereinstimmung der Wirkungsrichtung der Faktoren für Verständlichkeit und Natürlichkeit spiegelt sich in einer positiven Korrelation der beiden Ratings wider (Pearson's Produkt-Moment-Korrelation:  $r = 0.512$ , mit  $t = 22.9988_{df\ 1486}$ ,  $p < 0.0001$ ). Dies trägt allerdings nicht viel zur Klärung ihres Zusammenhangs bei, denn die Verständlichkeit könnte die Wahrnehmung der Natürlichkeit bedingen oder aber gleichermaßen die Natürlichkeit die Verständlichkeit beeinflussen. Auch eine mangelnde Diskriminierung der

<sup>42</sup>Linear mixed model fit by REML:

Formula:  $z.rating \sim audio * experiment + log.latency + (1 + item.pos | subject) + (1 | item)$

Der Slope für die Item-Position (item.pos) in Abhängigkeit des VTn (subject) wurde aufgenommen, um die vom jeweiligen Versuchsteilnehmer abhängigen Gewöhnungseffekte auf das Rating durch die Ähnlichkeit der Stimuli zu berücksichtigen.

<sup>43</sup>Linear mixed model fit by REML:

Formula:  $z.rating \sim audio * experiment + log.latency + (1 + item.pos | subject) + (1 | item)$

Aufgaben durch die VTn könnte dies bewirkt haben. Denkbar ist zudem eine Beeinflussung der beiden Ratings durch die alleinige Tatsache ihrer sukzessiven Abfolge.

Die Ergebnisse der logarithmierten Latenzen spiegeln teilweise die Ergebnisse der Ratings wider. Eine nähere Erläuterung der Befunde für die Latenzen wollen wir an dieser Stelle aber aussparen, da sie nichts zur Diskussion des Sachverhalts beiträgt.<sup>44</sup>

#### 7.3.6.3. Diskussion: Einfluss der Prosodie

Die Ergebnisse bestätigen unsere Vermutung viel deutlicher als angenommen. Offenbar wirken sich die Unterschiede durch das Einschneiden der LSK sehr wohl negativ auf die Wahrnehmung der Prosodie aus, wobei unklar bleibt, ob den VTn die Manipulation bewusst wurde. Offenbar schlägt sich die verminderte Natürlichkeit der Prosodie auch direkt in der Verständlichkeit der Sätze nieder. In dieser Hinsicht müssen wir diese Unnatürlichkeit in unseren Ergebnisdiskussionen berücksichtigen, auch wenn die Entscheidung zum manipulierten Audiomaterial ganz bewusst und aus gutem Grund getroffen wurde.

Allerdings sind die Ergebnisse in verschiedener Hinsicht zu relativieren: (a) Wir vergleichen in der Normierungsstudie kontrastierte Varianten (geschnittene vs. ungeschnittene Sätze). Demgegenüber verwenden wir in Experiment 2 nur eine Variante (ausschließlich geschnittene Sätze). Es ist anzunehmen, dass die Manipulation ohne den Kontrast der ungeschnittenen Varianten nicht in dem Ausmaß auffallen würde wie in der Normierungsstudie. (b) In der Normierungsstudie wurde die Natürlichkeit der Prosodie explizit abgefragt und damit die Aufmerksamkeit der VTn direkt auf diesen Aspekt der Materialien gelenkt. In Experiment 2 wurde die Prosodie nicht thematisiert, zudem war als begleitende Aufgabe die Beobachtung des Bildschirms gegeben. Das ist ein weiterer Grund für die Annahme, dass die Manipulation während des Experiments nicht (oder zumindest kaum) bemerkt wurde. (c) Des Weiteren war zuerst die Verständlichkeit und anschließend die Natürlichkeit zu bewerten. Diese Reihenfolge war möglicherweise nicht ganz glücklich gewählt, denn sie begünstigt, dass sich eine negative Beurteilung der Verständlichkeit auf die Beurteilung der Prosodie niederschlägt. (d) Unser zentraler Experimentalfaktor in Experiment 2 (LSK vs. Kontrolle) beruht auf den gleichen geschnittenen Materialien: Die Kontroll-Sätze wurden auf Grundlage der Versionen mit eingeschnittener LSK durch zusätzliches Einschneiden eines Adverbs konstruiert. Folglich haben wir für den Faktor

---

<sup>44</sup>Auch für die Latenzen haben wir gemischte lineare Modelle gefittet. Während sich der Faktor Anapherentyp eher auf die Latenzen des Ratings auswirken, zeigt sich bei den Latenzen für die Ratings der Prosodie ausschließlich ein Effekt für den Audiotyp. Der Vollständigkeit halber seien die Modelle hier in verkürzter Weise mit den zentralen Befunden aufgeführt:

(a) Logarithmierte Latenzen beim Rating der Verständlichkeit: Linear mixed model fit by REML, Formula:  $\log.\text{latency} \sim \text{audio} * \text{experiment} + (1 + \text{item.pos} | \text{subject}) + (1 | \text{item})$ . Das Modell für die logarithmierte Latenz für die Bewertung der Verständlichkeit zeigt keinen Effekt für den Faktor Audiotyp ( $t = -2.14$ ,  $F = 1.0762$ ,  $p > 0.29$ ), aber signifikante Effekte für den Faktor Anapherentyp ( $t = -3.14$ ,  $F = 5.9832$ ,  $p < 0.05$ ) sowie für die Interaktion der beiden Faktoren ( $t = 1.99$ ,  $F = 3.9776$ ,  $p < 0.05$ ).

(b) Logarithmierte Latenzen beim Rating der Natürlichkeit der Prosodie: Linear mixed model fit by REML, Formula:  $\log.\text{latency} \sim \text{audio} * \text{experiment} + (1 + \text{item.pos} | \text{subject}) + (1 | \text{item})$ . Das Modell für die logarithmierte Latenz für die Bewertung der Natürlichkeit der Prosodie zeigt einen signifikanten Haupteffekt für den Faktor Audiotyp ( $t = -1.84$ ,  $F = 6.6355$ ,  $p < 0.05$ ), aber keinen Effekt für den Faktor Anapherentyp ( $t = -0.45$ ,  $F = 0.3582$ ,  $p > 0.54$ ) sowie keine Interaktion ( $t = 0.03$ ,  $F = 0.0009$ ,  $p > 0.9$ ).

Satztyp die gleichen Voraussetzungen. Die Effekte lassen sich also nicht alleine auf die Manipulation der Audio-Materialien zurückführen.

Ein interessanter Befund ist, dass sich der Faktor Anapherentyp in der geschnittenen und in der ungeschnittenen Variante sowohl auf die Verständlichkeit als auch auf die Wahrnehmung der Natürlichkeit ausgewirkt hat. Dies bestätigt zunächst einmal den Befund aus der Lesezeitmessung mit Self-Paced Reading und Verständnisfrage (Experiment 1) hinsichtlich des Aspekts der Verständlichkeit: Hier zeigte sich im Anteil korrekt beantworteter Fragen sowie in den Beantwortungslatenzen ein klarer Nachteil für die Variante mit Personalpronomen. Dieses Ergebnis deckt sich recht gut mit dem hier gefundenen Effekt der geringeren Verständlichkeit. Erstaunlich ist allerdings, dass die Version mit Personalpronomen nicht nur in der geschnittenen, sondern auch in der ungeschnittenen Variante ein schlechteres Rating für die Natürlichkeit erhielt. Dies scheint unverständlich, denn in der ungeschnittenen Variante sollte sich hierfür eigentlich kein Unterschied zeigen. Eine mögliche Erklärung bietet die oben erwähnte Auswirkung des ersten Ratings auf das zweite Rating. So könnte, neben der mangelnden Diskriminierung zwischen den beiden Ratings, die Beurteilung der Natürlichkeit der Prosodie tatsächlich stark von der jeweils zuvor getroffenen Beurteilung für die Verständlichkeit beeinflusst worden sein.

Trotz der diskutierten Einschränkungen bleibt festzuhalten, dass die Audio-Manipulation einen negativen Effekt auf die Wahrnehmung der Prosodie und auf die Verständlichkeit der Sätze hat. Vor allem die geringere Verständlichkeit könnte verstärkend Einfluss auf den LSK-Effekt genommen haben, bzw. der Effekt der LSK könnte durch die schlechtere Verständlichkeit begünstigt worden sein. Wir werden diesen Aspekt in der Abschlussdiskussion des Kapitels im Zusammenhang mit dem Begriff Shallow Parsing wieder aufgreifen, nach dem LSK-Effekte möglicherweise als Effekte flacher Analyse durch überlastetes oder zusammengebrochenes Parsing erklärt werden könnten.

## 7.4. AnaLoCo 3: Visual-World-Experiment mit Szenen

Im zweiten Visual-World-Experiment (Experiment 3) verfolgen wir im Grunde dieselbe Methodik, dasselbe Design und dieselben Fragen wie in Experiment 2. Allerdings haben wir verschiedene Gesichtspunkte deutlich verändert, um die Schwächen des vorausgegangenen Experiments aufzufangen und seine Einschränkungen zu überwinden. Die Veränderungen betreffen sowohl die visuellen als auch die sprachlichen Stimuli.<sup>45</sup>

Anstatt kontextfreie einzelne Referenten abzubilden, verwenden wir in Experiment 3 szenische Darstellungen, die die benannten Referenten bei der Ausführung der Handlungen abbilden, die durch die Verben bezeichnet werden. Es werden jeweils vier Abbildungen mit verschiedenen Handlungen und verschiedenem Agens zeitgleich präsentiert. Auf der

---

<sup>45</sup>Das neue Design für die Kontrolle der sprachlichen Stimuli und für die Umstellung auf szenische Abbildungen wurden von Lars Konieczny in Zusammenarbeit mit Sascha Wolfer und dem Autor entwickelt. Die Idee entstand aus Diskussionen der Ergebnisse und Probleme des ersten Visual-World-Experiments. Material und Design basieren folglich auf dem Material des Vorgängerexperiments. Die Erstellung bzw. Aufbereitung der Materialien und die Erhebung der Daten wurden im Rahmen eines Projektseminars an der Universität Freiburg unter der Betreuung von Lars Konieczny, Sascha Wolfer und dem Autor geleistet.

Position der Anapher entsprechen die Szenen der globalen, der lokalen oder aber keiner der beiden Interpretationen. Die Darstellung von Szenen soll das Problem der Unklarheit des determinierenden Faktors der Aufmerksamkeitssteuerung abfangen, das bei der Zuordnung von Einzelreferenten während der Satzverarbeitung entstand: Wird das aktuell genannte Objekt (bzw. das Antezedens) oder aber das aktuelle Agens der beschriebenen Handlung fixiert? Die abweichenden Möglichkeiten bestehen bei einer szenischen Abbildung nicht mehr, diese ermöglichen eine eindeutige Zuordnung der Reaktion. Wir erwarten unter den LSK-Bedingungen einen höheren Anteil an Fixationen auf der Szenenabbildung, die der Interpretation der LSK entspricht.

Den sprachlichen Stimuli legen wir das Material aus den Vorgängerexperimenten zugrunde, verwenden aber in den Kontrollbedingungen eine andere Form der Durchbrechung der LSK, die keine weitere mögliche LSK enthält und die keine unterschiedliche Satzlänge vor der Anapher bedingt. Das bedeutet, die Anapher folgt in allen vier Bedingungen auf der gleichen Position innerhalb eines Satzes. Durch die Vermeidung unterschiedlicher Satzlängen<sup>46</sup> vor der Anapher können wir sprachliche Störeffekte ausschließen, etwa die Menge intervenierenden Materials zwischen abhängigen Wörtern oder semantische Effekte durch die Einführung von Temporaladverbien. Vor allem aber vermeiden wir dadurch den Aufbau einer weiteren lokal kohärenten Subsequenz mit wiederum eigener Verteilung der Bindungsdomänen.

Mit diesem Vorgehen bei der Gestaltung der Stimuli gehen wir allerdings andere Kompromisse ein, da es seinerseits Einschränkungen für mögliche Aussagen aus den Ergebnissen birgt. Unter anderem wäre zu fragen, ob die szenischen Darstellungen Aussagen über die Verarbeitung der Anaphern zulassen. Zudem könnte ihr Vorhandensein zu einem Bias für mögliche Interpretationen zugunsten der LSK führen. Wir werden in der Diskussion darauf zu sprechen kommen und darlegen, warum sich die beiden Experimente bzw. ihre Befunde ergänzen und ihre Kombination Aussagen über die Verarbeitung anaphorischer Referenzen in LSK ermöglicht.

### 7.4.1. Design

Wir verwenden wie gehabt ein  $2 \times 2$ -Design mit den Faktoren Anapherotyp und Satztyp in den identischen Bedingungen aus Experiment 2. Tabelle 7.11 gibt die Kodierung der Bedingungen wieder. Diese werden in nachfolgender Liste aufgeführt.

Design		Faktor 1: Anapherotyp	
		sich	ihn
Faktor 2: LSK	mit LSK	1	2
	LSK durchbrochen	3	4

Tabelle 7.11.: Tabelle zum Design für Experiment 2.

#### **Bedingung 1** Reflexivpronomen (*sich*) mit lokal interferierender Lesart (pro LSK).

<sup>46</sup>Dem ist anzumerken, dass die Satzlängen in unserer Umsetzung tatsächlich nur hinsichtlich der Anzahl der Wörter gleich sind, nicht aber hinsichtlich der Silbenanzahl. Dennoch wurde damit die Satzlänge über die Bedingungen hinweg substanziell angeglichen.

**Bedingung 2** Personalpronomen (*ihn*) mit lokal interferierender Lesart (pro LSK).

**Bedingung 3** Reflexivpronomen (*sich*), unterdrückte lokale Lesart (contra LSK).

**Bedingung 4** Personalpronomen (*ihn*), unterdrückte lokale Lesart (contra LSK).

#### 7.4.2. Stimulus-Materialien

**Sprachlich-auditive Stimuli** Für die Satzmaterialien wurde eine Auswahl von 24 Sätzen aus dem zugrundeliegenden Material (48 Sätze) entnommen.<sup>47</sup> Die ausgewählten Sätze waren durch die in ihnen verwendeten Verben am besten für die geplanten Modifikationen geeignet. Die Sätze mit LSK entsprechen weiterhin exakt dem Aufbau aus dem Vorgängerexperiment: Sie bestehen aus einem Objekt-Relativsatz, der zentral in einen temporalen Nebensatz eingebettet ist. Nach dem Relativsatz folgt die anaphorische Referenz als Satz-Objekt des temporalen Nebensatzes sowie eine Präpositionalphrase und das finite Verb des Nebensatzes. Abschließend folgt ein Hauptsatz mit entsprechend topikalisiertem finitem Verb. Die beiden Versionen der Sätze mit Reflexiv- und Personalpronomen sind in den Beispielen 28a und 28b ersichtlich.

In den Kontrollbedingungen wird die gleiche Struktur aufrechterhalten, es wurde in dieser Studie kein Adverb eingefügt. Allerdings wurde das finite Verb des Relativsatzes durch ein finites Partikelverb ausgetauscht. Die Partikelverben haben dabei dieselben Eigenschaften wie die normalen Verben: Sie können sowohl reflexiv als auch transitiv verwendet werden und würden folglich mit einem Reflexivpronomen und gleichermaßen mit einem Personalpronomen einen semantisch korrekten Satz ergeben. Deutsche Partikelverben (z.B. *anziehen*, bestehend aus dem Verbstamm *ziehen* und einem Partikel, der Präposition *an*) haben die grammatikalische Eigenschaft, innerhalb von Hauptsätzen einen Verbrahmen aufzubauen, der aus dem finiten Verbstamm, dazwischenliegenden Satzargumenten, und dem abgetrennten Partikel besteht (z.B. *Der Vater zieht den Sohn an.*). In Nebensätzen hingegen werden Partikel und Verbstamm des Partikelverbs zusammengezogen (z.B. ..., *als der Vater den Sohn anzieht.*). Durch das Partikelverb am Ende des Relativsatzes, das auf dieser Position folglich in Nebensatz-Morphologie erscheint (d.h., zusammengezogen ist), wird die LSK in der Kontrollversion unterbunden, denn durch die morphologische Form des Verbs ist die lokal isolierte Hauptsatz-Lesart auch lokal ausgeschlossen. Die beiden Versionen der Kontrollsätze für die beiden Anapherntypen sind ersichtlich in Beispiel 28c und 28d.

Die gewählte Vorgehensweise bei der Erstellung der Kontrollbedingungen hat drei Vorteile: Erstens sind die Sätze über die Bedingungen hinweg gleich lang (gleiche Wortanzahl). Zweitens wird kein anderes lexikalisches Element eingeführt, das potentiell weiteren Verarbeitungsaufwand fordern könnte. Drittens wird mit diesem Kontrollverfahren keine weitere LSK-Struktur erzeugt, wie sie in Experiment 2 durch das eingefügte Adverb entstand. Beim Verbtasch haben wir in allen Sätzen jeweils zwei gleichbleibende Verben eingesetzt, die nur in ihren Positionen vertauscht werden: In den Sätzen in LSK-Bedingungen ist das Partikelverb auf der finalen Position des Nebensatzes (d.h. außerhalb der LSK), während

---

<sup>47</sup>Der elektronische Anhang (PDF-Format) enthält auch das vollständige sprachliche und visuelle Stimulus-Material des Experiments AnaLoCo 3. Der Anhang wird auf Anfrage per E-Mail (*helmut.weldle@googlemail.com*) zugesandt.

die kritische Position am Ende des Relativsatzes durch ein normales finites Verb belegt wird. In den Kontrollsätzen werden diese Verbpositionen vertauscht, so dass nun ein finites Partikelverb die finale Position im Relativsatz einnimmt. Durch dieses Vorgehen wird erreicht, dass bei den korrespondierenden Abbildungen zu den Sätzen die abzubildenden Handlungen über die Bedingungen gleich bleiben, und die Abbildungen über die verschiedenen Bedingungen somit besser kontrolliert sind.

- (28) Auf den folgenden Abbildungen sehen wir einen Vater und einen Sohn in vier Szenen.
- a. Als der Vater<sub>i</sub>, den der Sohn<sub>j</sub> kämmt, sich<sub>i/\*j</sub> im Wohnzimmer anzieht, ist die halbe Stadt bereits auf dem Weg zur Aufführung.
  - b. Als der Vater<sub>i</sub>, den der Sohn<sub>j</sub> kämmt, ihn<sub>\*i/j...</sub> im Wohnzimmer anzieht, ist die halbe Stadt bereits auf dem Weg zur Aufführung.
  - c. Als der Vater<sub>i</sub>, den der Sohn<sub>j</sub> anzieht, sich<sub>i/\*j</sub> im Wohnzimmer kämmt, ist die halbe Stadt bereits auf dem Weg zur Aufführung.
  - d. Als der Vater<sub>i</sub>, den der Sohn<sub>j</sub> anzieht, ihn<sub>\*i/j...</sub> im Wohnzimmer kämmt, ist die halbe Stadt bereits auf dem Weg zur Aufführung.

Für die Präsentation wurden wieder für alle 96 Sätze (24 Items in vier Bedingungen) auditive Versionen erstellt. Dazu wurde der Basissatz in normalem Sprechtempo aufgezeichnet.<sup>48</sup> Wie auch in der ersten Visual-World-Studie haben wir versucht, die Effekte prosodischer Strukturindikatoren durch ein Splicing-Verfahren zu kontrollieren, und damit ausschließlich syntaktische Informationen aufrechtzuerhalten. Wieder wurde hier die LSK als eigenständiger Hauptsatz aufgezeichnet, isoliert und in den Basissatz eingeschnitten. Dasselbe Vorgehen wurde für die Kontrollversion verwendet: Die Version mit Partikelverb wurde eingesprochen, als sei sie ein eigenständiger Hauptsatz mit korrekter Struktur.<sup>49</sup> In die beiden Versionen mit eingeschnittener Subsequenz wurde noch zur Erzeugung der Bedingungen mit Personalpronomen eben dieses an die Position des Reflexivpronomens eingesetzt. Außer den Ziel-Sätzen wurde zu jedem Item ein Einleitungssatz eingesprochen, der die beiden explizit genannten Personen vorstellt. Im Unterschied zu Experiment 2 wird hier keine dritte Person mehr eingeführt und auch nicht mehr abgebildet. Die Reihenfolge der beiden Personen in der Einleitung wurde über die Sätze rotiert, damit daraus nicht auf die Benennungsreihenfolge im Ziel-Satz geschlossen werden kann.

Für die Synchronisation der kritischen Positionen in der Analyse wurden wiederum für jeden Satz die Zeitmaße aufgezeichnet, die wir auch im vorangegangenen Experiment verwendet haben: (a) Die Gesamtlänge des jeweiligen Satzes, (b) der On-Set des Pronomens, (c) der Off-Set der LSK. Für die Darbietung wurden die Sätze nach der Latin-Square-Methode vollständig rotiert.

Das Experiment wurde parallel zu einem weiteren Visual-World-Experiment durchgeführt, dessen Items als Füllmaterial dienten. Dieses umfasste 24 Experimentalsätze

<sup>48</sup>Das Satzmaterial wurde wiederum von nur einer Person eingesprochen.

<sup>49</sup>Dabei wurde das Partikelverb so gesprochen, als sei es ein normales Verb. Dies führt natürlich zur Möglichkeit, dass der Sprecher diese unnatürliche Satzform prosodisch auffällig markiert. Mangels besserer Kontrollmöglichkeiten müssen wir uns hier auf das subjektive Sprach- und Hörgefühl der Projektleiter und -teilnehmer verlassen, die sich darin einig waren, dass es hinsichtlich der Prosodie keine Unterschiede oder Auffälligkeiten gab.

in vier Bedingungen. Das Füllmaterial hatte eine vergleichbare Komplexität zu unseren Experimental-Items. Zudem wurde dasselbe visuelle Präsentationsschema für die Darbietung der visuellen Materialien verwendet. Insgesamt ergibt sich daraus ein Verhältnis von 1:1 (50%) zwischen Experimental-Items und Füll-Items.

**Visuelle Stimuli** Der Aufbau der visuellen Stimuli bzw. ihr Bezug zu den Satzbedingungen ist recht komplex. Zugrunde liegt die Idee, immer jeweils eine Szenenabbildung für die globale Interpretation und die Interpretation der LSK darzubieten. Für eine tatsächliche Verarbeitung der LSK wird in den entsprechenden Bedingungen eine Verschiebung der relativen Anzahl der Fixationen im Vergleich zur Kontrollbedingung auf die Abbildung der Handlung der LSK erwartet.

Für jede Bedingung werden immer vier Abbildungen dargeboten, die in einem Rechteck auf dem Bildschirm angeordnet sind. In den Bedingungen mit Reflexivpronomen werden nur reflexive Handlungen abgebildet, d.h. es wird jeweils eine Person abgebildet, die eine Handlung mit sich selbst vollzieht. In den Versionen mit Personalpronomen hingegen werden nur transitive Handlungen abgebildet, d.h. es werden immer zwei Personen abgebildet, wobei Person A eine Handlung ausführt, die auf Person B gerichtet ist. Die abgebildeten Personen sind immer die explizit benannten Personen aus dem temporalen Nebensatz und dem Relativsatz. Die abgebildeten Handlungen sind die durch die beiden Verben aus dem temporalen Nebensatz und dem Relativsatz benannten Handlungen. Für die Abbildungen werden die beiden Personen und die beiden Handlungen vollständig miteinander kombiniert.

Wir werden die Kombinationen für die reflexiven und die transitiven Abbildungen am Beispiel erläutern. In unserem Beispielsatz verwenden wir die beiden Personen *Vater* und *Sohn* sowie die beiden Handlungen *anziehen* und *kämmen*. Durch die vollständige Kombination ergeben sich im reflexiven Fall (Satzbedingungen 1 und 3) folgende vier Szenen (vgl. Abbildung 7.13): (a) der VATER KÄMMT SICH, (b) der SOHN KÄMMT SICH, (c) der VATER ZIEHT SICH AN, (d) der SOHN ZIEHT SICH AN. Die jeweils nicht handelnde Person wird in neutraler Position daneben stehend abgebildet. In der Satzbedingung mit LSK entspricht Abbildung (c) der globalen Interpretation und Abbildung (b) der lokalen Interpretation der LSK. Die Abbildungen (a) und (d) entsprechen keiner global oder lokal möglichen Interpretation des Satzes.

Für die transitiven Bedingungen (Satzbedingungen 2 und 4) ergeben sich bei vollständiger Kombination die folgenden Szenen (vgl. Abbildung 7.14): (a) der VATER KÄMMT DEN SOHN, (b) der SOHN KÄMMT DEN VATER, (c) der VATER ZIEHT DEN SOHN AN, (d) der SOHN ZIEHT DEN VATER AN. In der Satzbedingung mit LSK entsprechen die Abbildungen (b) und (c) der globalen Interpretation. Abbildung (b) entspricht zusätzlich der lokalen Interpretation der LSK. Die lokale und die globale Interpretation überschneiden sich also – ein Umstand, auf den wir in einem Abschnitt zu den Hypothesen eingehen werden. Die Abbildungen (a) und (d) entsprechen wiederum keiner möglichen Interpretation des Satzes.

Die jeweilige Darbietung aller kombinatorischen Möglichkeiten erfüllt verschiedene Zwecke. Zum einen ist die Abbildung beider durch die Verben beschriebenen Aktionen nötig, da die Position des Verbs vertauscht wird. Somit tauchen in jeder der vier Bedingungen alle Handlungen in der transitiven wie der reflexiven Lesart auf. Zum anderen soll

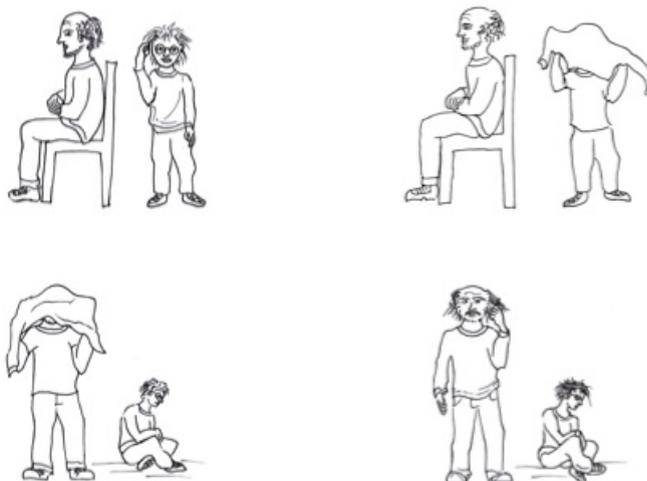


Abbildung 7.13.: Visuelle Stimuli im zweiten Visual-World-Experiment: Abbildung reflexiver Szenen.



Abbildung 7.14.: Visuelle Stimuli im zweiten Visual-World-Experiment: Abbildung transitiver Szenen.

die Abbildung jeweils aller vier Szenen verhindern, dass die VTn die zugrundeliegende Idee und den Gegenstand der Untersuchung erkennen können. Damit zusammenhängend soll die Abbildung aller vier Szenen auch der Verwendung von Strategien durch den VT vorbeugen, denn über mehrere Sätze betrachtet ist die Möglichkeit nicht auszuschließen, dass die Systematik der Beziehung zwischen den Sätzen und den Abbildungen zumindest oberflächlich entdeckt wird.<sup>50</sup>

Die Einzelszenen wurden in der Größe 650 x 450 Pixel (bei einer Gesamtgröße und Auflösung des Bildschirms von 1600 x 1200 Pixel) in einer rechteckigen Anordnung abgebildet. Die Reihenfolge der Szenen im Rechteck wurden komplett nach der Latin-Square-Methode durchrotiert, um einen Positionseffekt auszuschließen.

### 7.4.3. Methodik und Hypothesen

Das Visual-World-Experiment 3 entspricht dem Vorgehen aus Experiment 2, durch die Verwendung szenischer Abbildungen ist sie aber methodisch näher an den Studien von Konieczny und Müller (2007); Konieczny, Müller, Hachmann et al. (2009). Zur Aufzeichnung der Fixationen wurden nicht-überlappende IAs definiert, die in ihrer Ausdehnung der Größe der Einzelszenen-Abbildungen entsprechen. Die hierauf gerichteten Fixationen wurden in ihrer zeitlichen Auflösung auf die Zeitsignatur des dargebotenen Satzes abgebildet. Zur Implementierung des Experiments wurde wiederum die Software Experiment Builder und zur Durchführung und Aufzeichnung das Remote-Eyetracker-System Eyelink 1000 von SR Research (ohne Verwendung einer Kopfstütze) verwendet.

#### 7.4.3.1. Ablauf und Teilnehmer

Am Experiment nahmen 36 VTn teil, vorwiegend Studierende der Universität Freiburg. Die Teilnehmer erhielten im Anschluss an das Experiment einen Versuchsteilnehmerstundenschein oder wurden gemäß dem Standardsatz finanziell entlohnt (vgl. Kap. 7.3.3.1).

Vor dem Experiment wurden in Papierform Instruktionen zum Ablauf und zur Aufgabe gegeben. Nach dem Lesen der Instruktionen konnten die VTn Verständnisfragen stellen. Das Experiment wurde parallel zu einem weiteren Visual-World-Experiment durchgeführt, dessen Items wie im Abschnitt zu den Materialien beschrieben als Füllmaterial dienten. Nach zwei Übungsdurchgängen folgten die Experimental-Items. Nach dem Experimenten mussten die VTn einen Fragebogen zu ihren Lesegewohnheiten und ihrem Sprach- und Schriftspracherwerb ausfüllen.

Wie in Experiment 2 war die Primär-Aufgabe der VTn, sich die Experimentalsätze anzuhören, zu versuchen diese zu verstehen und parallel dazu auf dem Bildschirm die Abbildungen der Szenen zu betrachten. Während des Experiments starteten die Items jeweils durch Drücken einer Tastaturtaste (hier: Leertaste). Hierdurch wurde die jeweilige Abbildung mit allen vier Szenen eingeblendet und blieb bis zum Abschluss des Trials permanent eingeblendet. Unmittelbar nach Einblendung der Szenen wurde per Lautsprecher das sprachliche Stimulusmaterial präsentiert, nach dem beschreibenden Einleitungssatz

---

<sup>50</sup>So könnte erkannt werden, dass immer die erstgenannte Person eine reflexive Handlung und die immer die zweitgenannte Person eine transitive Handlung ausführt.

folgte der jeweilige Zielsatz. Sekundär-Aufgabe war es, nach Abschluss des jeweiligen Zielsatzes (markiert durch einen Glockenton unmittelbar nach Ende des Satzes) mit der Computermaus aus den vier abgebildeten Szenen diejenige anzuklicken, die tatsächlich durch den gesprochenen Satz beschrieben wurde. Im Unterschied zu Experiment 2 gab es in dieser Version korrekte und falsche Antworten. Nach dem Anklicken der Szene wurde eine Rückmeldung eingeblendet, ob die Auswahl richtig oder falsch war. Die zusätzliche Aufgabe sollte einerseits sicherstellen, dass die VTn dem gesprochenen Material inhaltlich folgen und einen Bezug zwischen dem gehörten Satz und den abgebildeten Szenen herstellen. Andererseits wurde sie aufgezeichnet, um anschließend ein hinreichendes Verständnis der Sätze zu beurteilen.

### 7.4.3.2. Hypothesen

Unter der Annahme, dass sich die Fixationen auf das Agens der zuletzt beschriebenen Handlung beziehen, sollte sich durch die Verwendung von szenischen Darstellungen mit eindeutig identifizierbarem Agens zeigen, welche Lesart zu einem bestimmten Zeitpunkt präferiert wird. Wir nehmen an, dass sich lokale Interferenzen auf die Auflösung der anaphorischen Referenz und damit unmittelbar auf die Interpretation des sprachlich beschriebenen Ereignisses auswirken. In der Folge sollten sich temporär Fixationen auf Situationsabbildungen finden, die von der globalen Lesart abweichen. Auf dieser Grundlage ist zu erwarten, dass die Ergebnisse der Fixationen auf der kritischen Region eindeutig für oder gegen lokale Kohärenzen sprechen und sich die Ergebnisse aus dieser neuen Perspektive szenen-basierter visueller Stimuli klar deuten lassen.

Wir erwarten wir für die beiden Anapherntypen folgende Effekte hinsichtlich des zeitlichen Verlaufs ihrer Verarbeitung, des Wirkens der experimentellen Bedingungen sowie der Verschiebung der Anteile der visuellen Aufmerksamkeit:

#### 1. Auswirkungen und Einsatz der Effekte für das Reflexivpronomen.

- Wir erwarten für die Bedingung LSK im Vergleich zur Kontrollbedingung eine Verschiebung der relativen Häufigkeit der Fixationen von der Abbildung mit dem Subjekt des Matrixsatzes als Agens (globales Agens, Subjekt des temporalen Nebensatzes) auf die Abbildung mit dem Subjekt des Relativsatzes als Agens (lokales Agens, Subjekt des Relativsatzes). Im Beispielsatz entspricht dies der Verschiebung der Fixationswahrscheinlichkeit von der Abbildung, die den VATER als Agens der reflexiven Handlung zeigt, hin zur Abbildung, die den SOHN als Agens der reflexiven Handlung zeigt. Dies entspricht einer Lesart im Sinne einer lokal isolierten Anwendung des Prinzips A entgegen der globalen Anwendung des strukturellen Bindungsprinzips.
- Folglich erwarten wir in der LSK-Bedingung (Bedingung 1, Bsp. 28a, LSK: *...der Sohn kämmt sich...*) einen höheren Gesamtanteil der Fixationen auf dem Bild mit lokalem Agens (Abb. 7.13, Szene links oben) als in der Kontrollbedingung (Bedingung 3, Bsp. 28c, LSK: *...der Sohn anzieht sich...*) auf dem korrespondierenden Bild (Abb. 7.13, Szene rechts oben).
- Gleichzeitig erwarten wir in der LSK-Bedingung (Bedingung 1) eine Verringerung des Gesamtanteils der Fixationen für das Bild mit dem globalen Agens

(Abb. 7.13, Szene links unten) als in der Kontrollbedingung (Bedingung 3) für das korrespondierende Bild (Abb. 7.13, Szene rechts unten).

- Wie bereits in der ersten Visual-World-Studie angemerkt, ist es schwierig, eine konkrete zeitliche Auflösung der Effekte vorherzusagen. Hinsichtlich der starken strukturellen Determiniertheit von Reflexivpronomen und in Relation zur zeitlichen Auflösung, die in der ersten Studie gefundenen wurde, erwarten wir einen frühen Effekt, der ähnlich wie in Experiment 2 ab etwa 500 ms einsetzen sollte.

## 2. Auswirkungen und Einsatz der Effekte für das Personalpronomen.

- Die Erwartungen der Effekte auf die Fixationen sind im Fall der Personalpronomen gänzlich anderer Natur als die bisher besprochenen Effekte, da die Interpretation der LSK mit der Interpretation nach globaler Lesart übereinstimmt. Zur Verdeutlichung: In der LSK-Bedingung (Bedingung 2, Bsp. 28b) haben wir zum einen die global korrekte Lesart *...den der Sohn kämmt...*, zum anderen die lokale Lesart *...der Sohn kämmt ihn...* – beide Lesarten korrespondieren mit der Abbildung, die den SOHN als Agens der KÄMMEN-Handlung zeigt (Abb. 7.14, Szene links oben).
- Allerdings korrespondiert die global korrekte Lesart das Personalpronomen betreffend (*Als der Vater ... ihn im Wohnzimmer anzieht...*) zur Abbildung, die den VATER als Agens darstellt (Abb. 7.14, Szene rechts unten). Die Frage ist also: Wie früh nehmen die Fixationen auf der Abbildung der lokalen Lesart zu Gunsten der Abbildung der ausschließlich globalen Lesart ab?
- Wir erwarten, dass die Abnahme im Fall der Kontrollbedingung früher anfängt bzw. im Fall der LSK-Bedingung später einsetzt, da durch die Kontrollstruktur die lokale Lesart durchbrochen und nicht über die Relativsatz-Grenze hinaus fortgesetzt werden kann. Dies könnte sozusagen als Ausdehnungs-Effekt der LSK zu Gunsten der lokalen Lesart bezeichnet werden. Dies sollte sich darin äußern, dass wir in der LSK-Bedingung einen höheren Anteil der Fixationen auf der Abbildung der lokalen Lesart erhalten (Abb. 7.14, Szene links oben) als für die Kontrollbedingung (*...den der Sohn anzieht...* bzw. *...der Sohn anzieht ihn...*) für die korrespondierende Abbildung (Abb. 7.14, Szene rechts oben).
- Auf wessen Kosten die erhöhten Fixationsanteile gehen, ist schwer abzuschätzen, da auf der Position der Anapher noch keine zweite Handlung benannt<sup>51</sup> und die zur globalen Lesart korrespondierende Abbildung folglich schwer zu identifizieren ist.
- Die Frage des Zeitpunkts ist für dieses Experiment eine andere als für das vorausgegangene Experiment 2: In der lokalen und zugleich globalen Lesart ist der *Vater* das Patiens der Handlung des Relativsatzes, die mit seiner Bindung an das Personalpronomen nur fortgesetzt werden würde. Die Frage lautet also vielmehr: Wie schnell wird erkannt, dass der *Vater* nicht Antezedens des Personalpronomens sein kann und eine neue Rolle einnehmen muss? Dies wird

---

<sup>51</sup>Das Verb des Matrixsatzes folgt erst am Satzende, da der Matrixsatz ein vorangestellter Nebensatz mit Letztstellung des finiten Verbs ist.

möglicherweise vor allem durch das Erkennen der Satzgrenze bestimmt und sollte in diesem Fall ebenfalls frühzeitig stattfinden.

## 7.4.4. Ergebnisse

### 7.4.4.1. Fixationsanteile auf den IAs

Über das Eyelink-System wurden die Fixationen auf den vier IAs aufgezeichnet. In einer ersten Analyse vergleichen wir wieder die Fixationsanteile auf den IAs getrennt für die Bedingung des Anapherotyps. Die Sekundär-Aufgabe ging nicht in die Analysen ein. Die Fixationen wurden dazu 100-ms-Zeitscheiben zugewiesen. Anschließend wurde innerhalb der Zeitscheiben der relative Anteil der Fixationen für die vier IAs berechnet. Die Synchronisation des zeitlichen Verlaufs über die Items erfolgt am On-Set des anaphorischen Ausdrucks (in den Diagrammen die gestrichelte vertikale Linie auf Zeitachsenabschnitt 0). Die Zeitscheiben wurden ab 2500 ms vor und bis 4000 ms nach dem Synchronisationspunkt berechnet.

Wir haben die Fixationsanteile auf den IAs wiederum in zwei Schaubildern getrennt nach Anapherotyp (Reflexivpronomen vs. Personalpronomen) abgetragen. Die Linien zeigen das arithmetische Mittel der Fixationsanteile auf je einer Interest Area in Abhängigkeit der Satzbedingung (mit oder ohne LSK) ab, das transparente Feld um den Mittelwert repräsentiert den Standardfehler. Der Linientyp unterscheidet die Satzbedingung, die den Fixationen zugrundeliegt:

- Die durchgängige Linie (mit gefüllten Mittelwert-Markierungen der Zeitscheiben) entspricht dabei jeweils der Bedingungen mit LSK (*...der Sohn kämmt sich...* und *...der Sohn kämmt ihn...*)
- Die gestrichelte Linie (mit leeren Mittelwert-Markierungen der Zeitscheiben) der Kontrollbedingung (*...der Sohn anzieht sich...* und *...der Sohn anzieht ihn...*)

Die Linienfarbe (sowie die Markierung der Zeitscheibe) unterscheiden die IAs in Abhängigkeit von der Experimentalbedingung:<sup>52</sup>

- Die schwarze Linie (Mittelwert-Markierungen der Zeitscheiben: Quadrate) repräsentiert die Fixationsanteile auf der Szene, die zur Handlung des Matrixsatz-Verbs korrespondiert. Dies ist im Fall der Sätze mit Reflexivpronomen die globale Lesart, im Fall der Sätze mit Personalpronomen die ausschließlich globale Lesart
- Die hellgraue Linie (Mittelwert-Markierungen der Zeitscheiben: Rauten) repräsentiert die Fixationsanteile auf der Szene, die zur Handlung des Relativsatz-Verbs

---

<sup>52</sup>Wir verzichten bei den folgenden Abbildungen auf die Darstellung der Fixationsanteile auf den IAs, die die jeweils nicht im Satz beschriebenen Situationen darstellen. Als Beispiel: In Satz 28a werden die Handlungen SOHN KÄMMT und VATERS ZIEHT AN beschrieben. Nicht beschrieben werden aber die Handlungen SOHN ZIEHT AN und VATERS KÄMMT. Diese IAs sind aber dennoch in die entsprechende Abbildung involviert, da wir ja immer die LSK-Bedingung mit der Kontrollbedingung vergleichen (28c), in der die jeweils umgekehrt verteilten Handlungen beschrieben werden. Wir kollabieren also sozusagen die IAs über die Bedingungen und erhalten deshalb nicht vier Linien (eine Linie pro Interest Area), sondern nur zwei Linien (eine Linie pro Bedingung) – und verbessern dadurch die Nachvollziehbarkeit der Diagramme wesentlich.

korrespondiert. Im Fall des Reflexivpronomens ist dies die lokale Lesart, im Fall des Personalpronomens die lokale und temporär globale Lesart

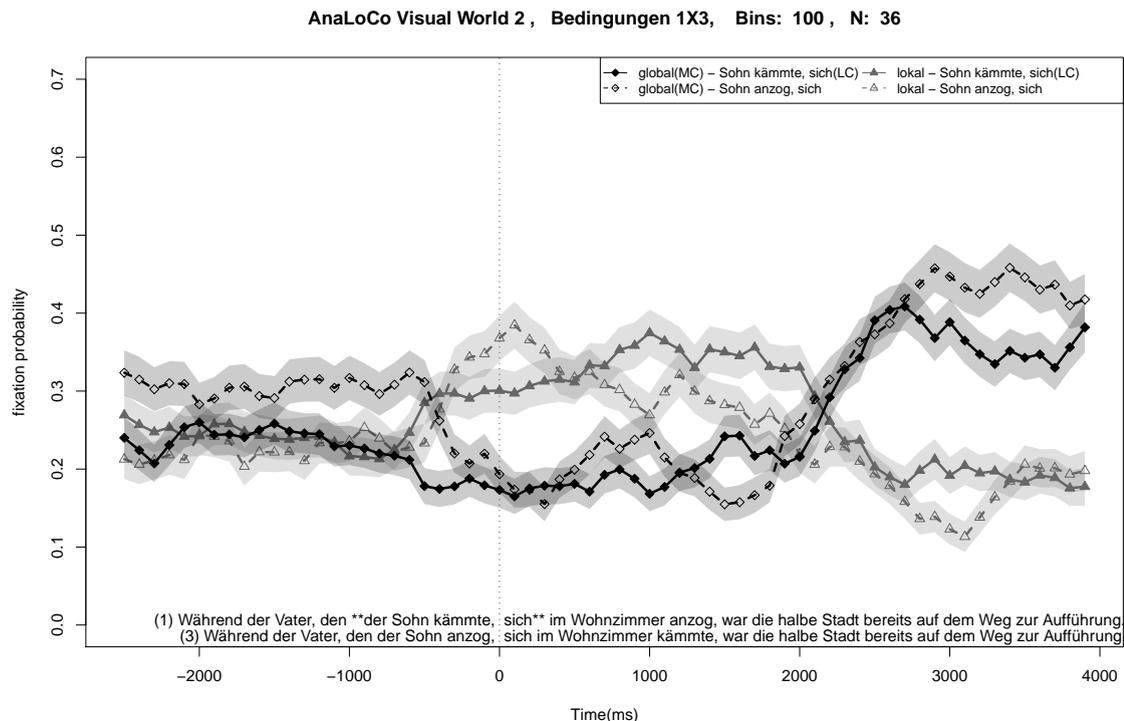


Abbildung 7.15.: AnaLoCo 3. Zeitlicher Verlauf der Fixationsanteile auf den IAs in Abhängigkeit der Satzbedingung (Reihenfolge der Verben) für die Bedingungen LSK vs. Kontrolle in den Versionen mit Reflexivpronomen. Relevant sind die Fixationsanteile nach dem Synchronisationszeitpunkt.

Diagramm 7.15 zeigt die gemittelte Fixationswahrscheinlichkeit für die reflexive Bedingung. Hier finden wir für die Bedingung mit LSK eine signifikant höhere Fixationswahrscheinlichkeit auf der Szene mit dem handelnden Subjekt des Relativsatzes (lokale Lesart). Dies indiziert, dass in dieser Bedingung die Interpretation der Subsequenz (...*der Sohn kämmt sich*...) verstärkt in Betracht gezogen wird. Der *Sohn* wird also entgegen der Komplementaritätsannahme als lokal zugängliches Antezedens des Reflexivpronomens angenommen. Auffällig ist, dass der Effekt in Experiment 3 deutlich später eintritt als in Experiment 2. In letzterem konnte der Effekt mit einem On-Set ab 500 ms festgestellt werden. In Experiment 3 hingegen erscheint der Effekt erst ab ca. 800 ms und zieht sich, abgesehen von einer kurzen zwischenzeitlichen Abschwächung, bis 2200 ms. Eine mögliche Erklärung dieser Verzögerung könnte die höhere Komplexität der visuellen Stimuli sein. Diese enthalten eine höhere Anzahl von Bildern, die zudem durch die Abbildung von Szenen komplexer sind als die kontextfreie Abbildung einzelner Referenten. Dies könnte zu einer längeren Suche nach der Szene führen, die zur aktuellen Interpretation passt.

Das zweite Diagramm (Abb. 7.16) zeigt die gemittelte Fixationswahrscheinlichkeit für die transitive Bedingung. Hier finden wir nun den Effekt, der in der ersten Visual-World-Studie (Experiment 2) ausblieb: Die Fixationswahrscheinlichkeit auf der Szene,

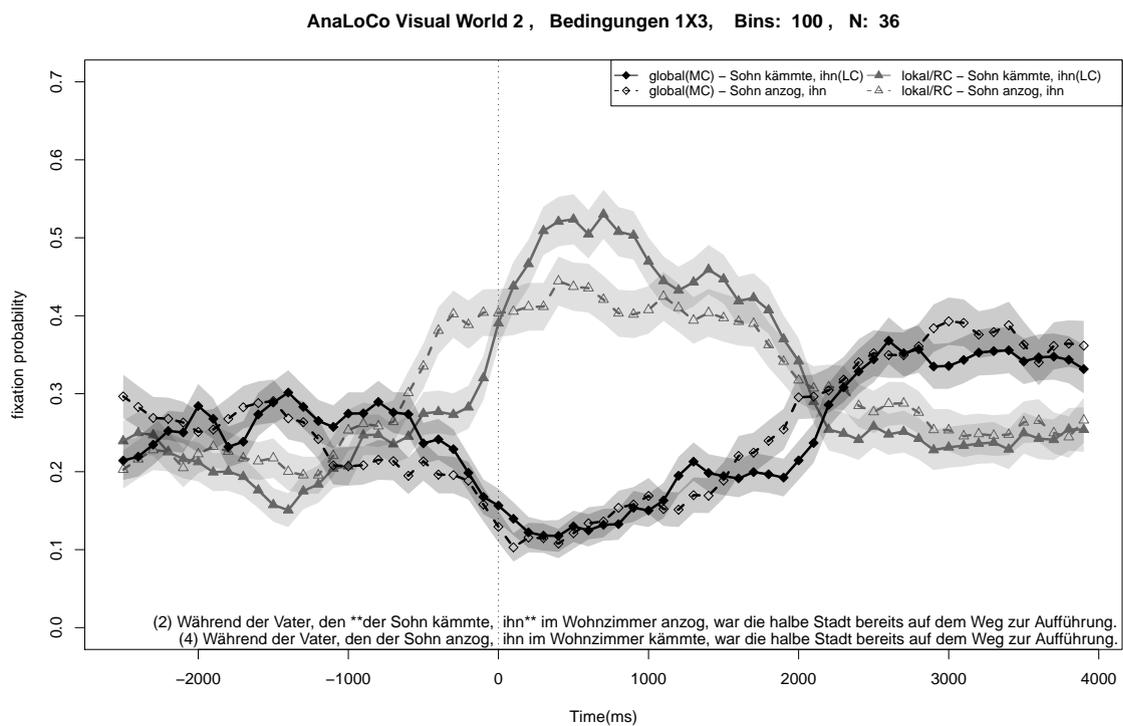


Abbildung 7.16.: AnaLoCo 3. Zeitlicher Verlauf der Fixationsanteile auf den IAs in Abhängigkeit der Satzbedingung (Reihenfolge der Verben) für die Bedingungen LSK vs. Kontrolle in den Versionen mit Personalpronomen. Relevant sind die Fixationsanteile nach dem Synchronisationszeitpunkt.

die das Subjekt des Relativsatzes beim Ausführen der transitiven Handlung zeigt, ist signifikant höher in der LSK-Bedingung. Es wird also länger die Abbildung der Handlung des Relativsatz-Verbs betrachtet bzw. es erfolgt ein späterer Wechsel zur Abbildung der ausschließlich globalen Lesart. Dies indiziert, dass der Vater – obwohl in der globalen Interpretation für die Bindung des Personalpronomens unzugänglich – länger die thematische Rolle des Patiens beibehält und folglich lokal als Antezedens des Personalpronomens akzeptiert wurde. Der Effekt setzt extrem früh ein und zieht sich von 200 ms bis 1000 ms nach On-Set der Anapher.

### 7.4.4.2. Gemischte lineare Modelle

Wir haben zur weiteren Überprüfung dieser Ergebnisse wiederum mit der Funktion *lmer* gemischte lineare Modelle für die Daten gefittet und dabei getrennte Modelle für die beiden Anapherarten berechnet. Wie in Experiment 2 verwenden wir als festen Faktor den Experimental-Faktor Satztyp hinsichtlich seiner Interaktion mit den drei IAs und als Zufallsfaktoren die Intercepts für VTn und Items. Abhängige Variable war wiederum das arithmetische Mittel der Fixationswahrscheinlichkeit für eine möglichst groß gewählte Zeitscheibe (800-1000 ms). Die Feststellung der Abgrenzung des zu untersuchenden Abschnitts konnte für Experiment 3 leider nicht so geradlinig vorgenommen werden wie in Experiment 2, denn die Auswahl des Startpunkts ab 500 ms nach On-Set der Anapher erwies sich als ungeeignet. Die Willkür der Wahl des On-Set wird in der Diskussion wieder aufgegriffen, wenn wir Gründe für vergleichsweise frühe oder späte Effekte suchen. Für die Bedingungen mit Personalpronomen haben wir als On-Set der Zeitscheibe ab 200 ms und für Reflexivpronomen ab 1000 ms nach dem On-Set der Anapher gewählt.

Wir stellen zunächst in Diagrammen das arithmetische Mittel und den Standardfehler für die gewählten Zeitscheiben in den beiden Bedingungen (vgl. Abb. 7.17 und 7.18) dar. Die darauffolgenden Tabellen (Tab. 7.12 und 7.13) enthalten die Daten der gemischten Modelle und die darin gefundenen Effekte sowie geschätzte F- und p-Werte.<sup>53</sup>

Für die Bedingung mit Reflexivpronomen zeigt sich wie erwartet eine hoch signifikante Interaktion zwischen Satz-Bedingung und Interest Area ( $F = 3.71$ ,  $p < 0.01$ ). Worauf ist diese Interaktion zurückzuführen? Beim Vergleich der t-Werte für die festen Faktoren des Modells liegt zunächst nahe, dass der Effekt nicht auf dem Unterschied für die Fixationen auf dem Ereignis der LSK (Agens und Handlung des Relativsatzes) beruht, sondern auf dem Unterschied auf der irrelevanten Interest Area nonmatch-Verb(RC). Diese zeigt das Agens des Matrixsatzes (VATER) als Ausführenden der reflexiven Handlung des Relativsatz-Verbs.

Ein Blick auf Diagramm 7.17 klärt diesen Sachverhalt: Zunächst einmal deutet sich hier in der Tat ein Effekt für die zweitgenannte Interest Area an – ein Umstand, auf den wir in

---

<sup>53</sup>Die Variablenamen in den gemischten Modellen entsprechen den Namen aus Experiment 2: mean\_POI = Fixationsanteile; Condition = Bedingungen des Faktors Satztyp; Condition-Kontrolle = Satzbedingung ohne LSK (Ausgangspunkt des statistischen Vergleichs); Condition-LSK = Satzbedingung mit LSK; IA = Interest Area; Item\_ID = Identität des Stimulussatzes; VTn = Identität des Versuchsteilnehmers. Zu den Variablen aus Experiment 2 kommen noch die Folgenden hinzu: IA-global = IA mit global korrekter Szene (Ausgangspunkt des statistischen Vergleichs); IA-lokal = IA mit Szene der LSK; IA-nonmatch-Verb(RC) = Szene für das Relativsatz-Verb mit nicht zum Satz korrespondierendem Agens; IA-nonmatch-Verb(MC) = Szene für das Matrixsatz-Verb mit nicht zum Satz korrespondierendem Agens; IA-unknown = Fixationen, die nicht zu einer IA zugeordnet werden können.

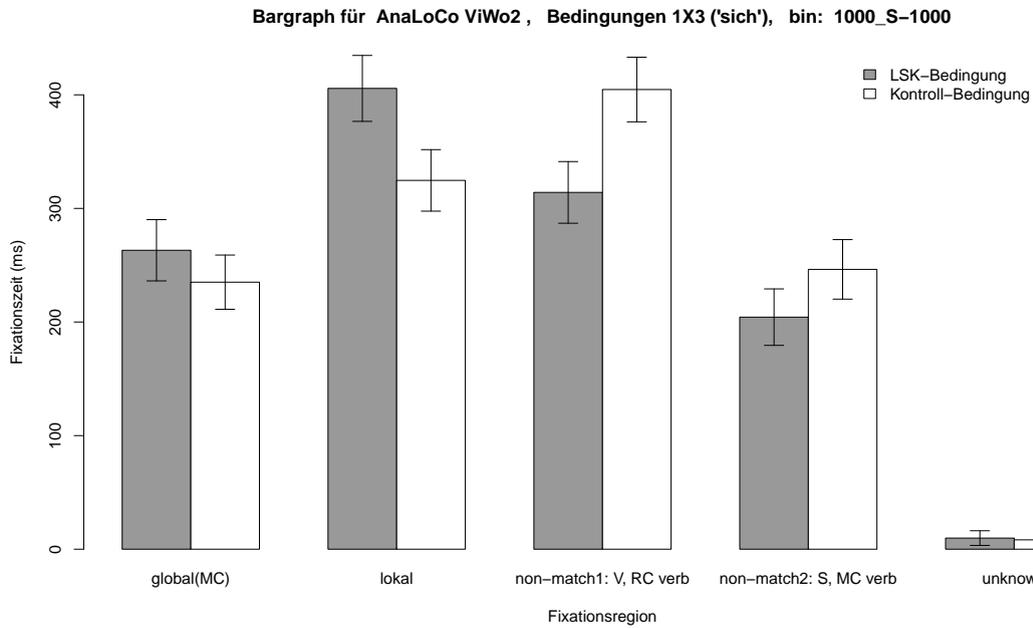


Abbildung 7.17.: AnaLoCo 3. Balkendiagramm für die Bedingung Reflexivpronomen: Fixationsanteile (y-Achse) für eine 1000 ms-Zeitscheibe (ab 1000 ms nach On-Set der Anapher) auf den vier IAs (x-Achse) in Abhängigkeit der Satzbedingung.

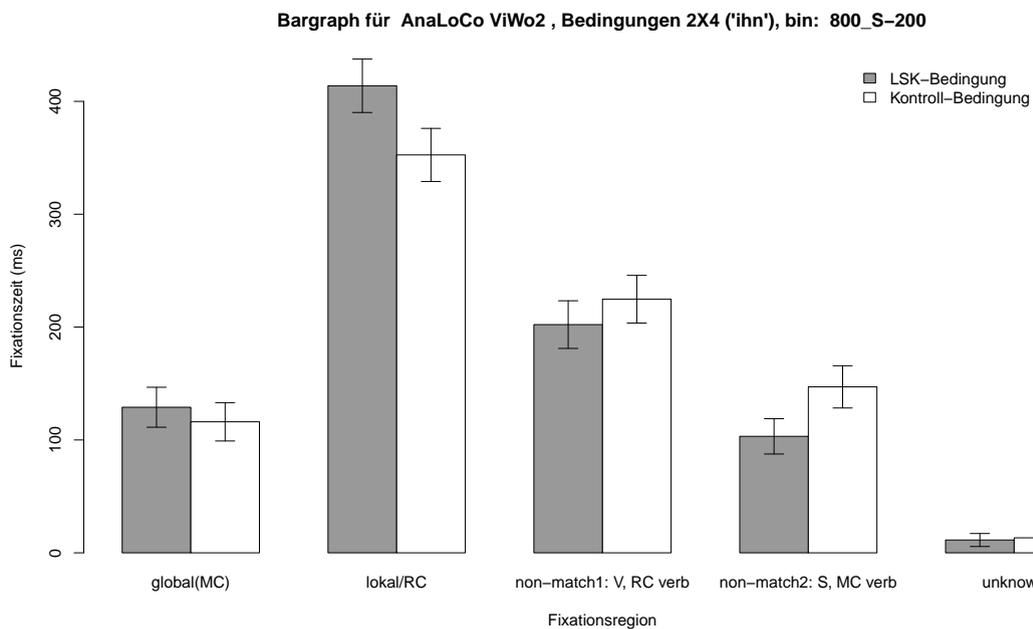


Abbildung 7.18.: AnaLoCo 3. Balkendiagramm für Bedingung Personalpronomen: Fixationsanteile (y-Achse) für eine 800 ms-Zeitscheibe (ab 200 ms nach On-Set der Anapher) auf den vier IAs (x-Achse) in Abhängigkeit der Satzbedingung.

Linear mixed model fit by REML

Formula: mean\_POI ~Condition \* IA + (1 | Item\_ID) + (1 | VTn)

Random effects:

<i>Groups</i>	<i>Name</i>	<i>Variance</i>	<i>Std.Dev.</i>
VTn	(Intercept)	0	0.00
Item_ID	(Intercept)	0	0.00
Residual		125107	353.71

Number of obs: 2160, groups: vp, 36; Item\_ID, 24

Fixed effects:

	<i>Estimate</i>	<i>Std. Error</i>	<i>t value</i>
(Intercept)	235.07	24.07	9.768
Condition-LSK	28.08	34.04	0.825
IA-lokal	89.59	34.04	2.632
IA-nonmatch-Verb(RC)	169.58	34.04	4.983
IA-nonmatch-Verb(MC)	11.26	34.04	0.331
IA-unknown	-226.85	34.04	-6.665
Condition-LSK:IA-lokal	52.89	48.13	1.099
Condition-LSK:IA-nonmatch-Verb(RC)	-118.64	48.13	-2.465
Condition-LSK:IA-nonmatch-Verb(MC)	-70.13	48.13	-1.457
Condition-LSK:IA-unknown	-26.53	48.13	-0.551

Analysis of Variance Table:

	<i>Df</i>	<i>Sum Sq</i>	<i>Mean Sq</i>	<i>F value</i>	<i>F</i>	<i>Df2</i>	<i>p</i>
Condition	1	10463	10463	0.0836	0.0836	2150	0.77246
IA	4	36098669	9024667	72.1354	72.1354	2150	0.00000
Condition:IA	4	1859794	464949	3.7164	3.7164	2150	0.00509

Tabelle 7.12.: AnaLoCo 3. Gemischtes lineares Modell für den Faktor Satztyp in der Bedingung Reflexivpronomen: Fixationsanteile auf den IAs für eine Zeitscheibe von 1000 ms ab einem On-Set 1000 ms nach dem Synchronisationspunkt.

der Diskussion näher eingehen werden. Zudem weisen die t-Werte auf einen signifikanten Haupteffekt mit einem höheren Anteil an den Fixationen für diese Interest Area noch vor dem Anteil der Fixationen auf der Interest Area der LSK hin. Gleichmaßen zeigt sich aber auch ein deutlicher Unterschied für die Interest Area mit Abbildung der LSK-Handlung. Da die Richtung dieses Unterschieds die gleiche Tendenz aufweist wie der Unterschied für die Abbildung der globalen Handlung, die in der Tabelle den Ausgangspunkt für den Interaktions-Vergleich stellt, wird der t-Wert für den erwarteten Unterschied nicht signifikant. Wie in Experiment 2 klärt ein weiteres Modell für den Faktor Satztyp isoliert für die Abbildungen der LSK (vgl. Tab. 7.14) den Sachverhalt auf und bestätigt einen signifikanten Haupteffekt für die Fixationsanteile auf den relevanten Abbildungen für die Bedingung ( $t = 2.061$ ,  $F = 4.2471$ ,  $p < 0.05$ ).

Linear mixed model fit by REML

Formula: mean\_POI ~Condition \* IA + (1 | Item\_ID) + (1 | VTn)

Random effects:

Groups	Name	Variance	Std.Dev.
VTn	(Intercept)	0	0.00
Item_ID	(Intercept)	0	0.00
Residual		70845	266.17

Number of obs: 2160, groups: vp, 36; Item\_ID, 24

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	116.06	18.11	6.409
Condition-LSK	12.91	25.61	0.504
IA-lokal/RC	236.47	25.61	9.233
IA-nonmatch-Verb(RC)	108.62	25.61	4.241
IA-nonmatch-Verb(MC)	30.97	25.61	1.209
IA-unknown	-102.78	25.61	-4.013
Condition-LSK:IA-lokal/RC	48.36	36.22	1.335
Condition-LSK:IA-nonmatch-Verb(RC)	-35.38	36.22	-0.977
Condition-LSK:IA-nonmatch-Verb(MC)	-56.75	36.22	-1.567
Condition-LSK:IA-unknown	-14.84	36.22	-0.410

Analysis of Variance Table:

	Df	Sum Sq	Mean Sq	F value	F	Df2	p
Condition	1	757	757	0.0107	0.0107	2150	0.91766
IA	4	33027521	8256880	116.5482	116.5482	2150	0.00000
Condition:IA	4	685148	171287	2.4178	2.4178	2150	0.04668

Tabelle 7.13.: AnaLoCo 3. Gemischtes lineares Modell für den Faktor Satztyp in der Bedingung Personalpronomen: Fixationsanteile auf den IAs für eine Zeitscheibe von 800 ms ab einem On-Set 200 ms nach dem Synchronisationspunkt.

Auch für die Bedingung mit Personalpronomen findet sich wie erwartet eine signifikante Interaktion zwischen Satz-Bedingung und Interest Area ( $F = 2.41$ ,  $p < 0.05$ ). Ein Vergleich der t-Werte legt aus dem gleichen Grund wie für die Reflexivpronomen nahe, dass dieser Unterschied vor allem auf dem Unterschied der Fixationsanteile für die irrelevante Interest Area nonmatch-Verb(MC) beruht. Diese zeigt das Agens des Relativsatzes (SOHN) als Ausführenden der transitiven Handlung des Matrixsatz-Verbs. Auch auf diesen Umstand werden wir in der Diskussion eingehen. Gleichmaßen deutet sich im Diagramm

7.18 aber wieder ein deutlicher Unterschied für die Interest Area mit LSK-Handlung an – auch hier mit gleicher Richtungstendenz wie für die Abbildung der globalen Handlung als Ausgangspunkt für den Vergleich der Interaktion. Ein weiteres Modell für den Faktor Satztyp isoliert für die Abbildungen der LSK (vgl. Tab. 7.15) klärt den Sachverhalt und bestätigt einen signifikanten Haupteffekt für die Fixationsanteile auf den relevanten Abbildungen für die Bedingung Satztyp ( $t = 1.966$ ,  $F = 3.8652$ ,  $p < 0.05$ ). Hinsichtlich der absoluten Verteilung der Fixationen weisen die t-Werte für den Faktor Interest Area auf einen deutlichen Unterschied der Verteilung hin (die ANOVA zeigt einen höchst signifikanten Haupteffekt mit  $F = 116.5482$ ,  $p < 0.0001$ ). Wie erwartet erhält die Interest Area, die die LSK und damit zugleich die global korrekte Interpretation der ersten Situation zeigt, den größten Anteil der Fixationen, gefolgt von der Interest Area nonmatch-Verb(RC).

Linear mixed model fit by REML

Formula: mean\_POI ~Condition + (1 | Item\_ID) + (1 | VTn)

Random effects:

Groups	Name	Variance	Std.Dev.
VTn	(Intercept)	3.7371e-06	1.9332e-03
Item_ID	(Intercept)	3.5499e+03	5.9581e+01
	Residual	1.6673e+05	4.0832e+02
Number of obs: 432, groups: vp, 36; Item_ID, 24			

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	325.20	30.35	10.716
Condition-LSK	80.97	39.29	2.061

Analysis of Variance Table:

	Df	Sum Sq	Mean Sq	F value	F	Df2	p
Condition	1	708102	708102	4.2471	4.2471	430	0.039918

Tabelle 7.14.: AnaLoCo 3. Gemischtes lineares Modell für den Faktor Satztyp in der Bedingung Reflexivpronomen: Fixationsanteile auf den IAs mit Abbildung der potentiellen LSK für eine Zeitscheibe von 1000 ms ab einem On-Set 1000 ms nach dem Synchronisationspunkt.

### 7.4.5. Diskussion AnaLoCo 3

Zunächst einmal bestätigen die Ergebnisse aus Experiment 3 die Befunde aus Experiment 2: Sowohl für die Bedingung mit Reflexivpronomen als auch für die Bedingung mit Personalpronomen zeigt sich ein Effekt der LSK. Wie erwartet zeigt sich der Effekt als vermehrte Aufmerksamkeit auf den IAs, die die LSK abbilden. Bei den Sätzen mit Personalpronomen entspricht dies einer vermehrten Fixation auf dem zur global korrekten Situation korrespondierenden Abbildung, für die die Aufmerksamkeit durch die LSK verstärkt bzw. verlängert wird. Dies könnte auch als eine spätere Suche nach einer nachfolgenden Handlung in der LSK-Bedingung gegenüber der Kontrollbedingung beschrieben werden.

Aber es gibt eine ganze Reihe offener Fragen, die wir hier beantworten wollen: (a) Unklar ist, warum der Effekt für die Reflexivpronomen später eintritt als dies in Experiment 2 der Fall war und von uns hinsichtlich der Bedingtheit durch strukturelle Constraints erwartet wurde. Noch nicht geklärt ist auch, warum sich der Effekt für die Versionen

Linear mixed model fit by REML

Formula: mean\_POI ~Condition + (1 | Item\_ID) + (1 | VTn)

Random effects:

Groups	Name	Variance	Std.Dev.
VTn	(Intercept)	12917.2	113.654
Item_ID	(Intercept)	2599.8	50.988
Residual		104871.0	323.838

Number of obs: 432, groups: vp, 36; Item\_ID, 24

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	351.68	30.88	11.387
Condition-LSK	61.26	31.16	1.966

Analysis of Variance Table:

	Df	Sum Sq	Mean Sq	F value	F	Df2	p
Condition	1	405353	405353	3.8652	3.8652	430	0.049939

Tabelle 7.15.: AnaLoCo 3. Gemischtes lineares Modell für den Faktor Satztyp in der Bedingung Personalpronomen: Fixationsanteile auf den IAs mit Abbildung der potentiellen LSK für eine Zeitscheibe von 800 ms ab einem On-Set 200 ms nach dem Synchronisationspunkt.

mit Personalpronomen so früh zeigt, da wir zumindest erwartet hätten, dass sich hier ein etwa zeitgleicher bzw. eher späterer Effekt zeigt. Während unsere Vorhersagen für den On-Set des Effekts in Experiment 2 also eingetroffen sind, zeigt sich in Experiment 3 ein völlig disparates Bild. (b) Unklar ist auch, warum die IAs nonmatch-Verb(RC) und nonmatch-Verb(MC), die die in unseren Hypothesen zunächst als irrelevant betrachteten wurden, für die Ergebnisse eine so große Rolle zu spielen scheinen. Denn die Zunahme der Fixationen auf der Abbildung der LSK-Situation führt nicht zu einer Verminderung der Fixationen auf der Abbildung der globalen Situation, sondern geht immer zu Lasten der Abbildung der IAs nonmatch-Verb(RC) und nonmatch-Verb(MC).

**Verschiebung der Fixationsanteile** Beginnen wir mit der Frage, wie und warum sich die Fixationsanteile bei den Reflexivpronomen verschieben: Die vermehrten Fixationen auf der LSK-Abbildung gehen zu Lasten der Abbildung nonmatch-Verb(RC). Aber warum wird die irrelevante Interest Area überhaupt betrachtet und warum wird gerade hiervon die Aufmerksamkeit abgezogen? Vergleichen wir dazu die beiden Abbildungen und den relevanten Satzausschnitt an einem Beispiel: Die LSK-Abbildung zeigt die Szene SOHN KÄMMT SICH, die nonmatch-Verb(RC)-Abbildung die Szene VATER KÄMMT SICH. Die Satzsequenz in der LSK-Bedingung entspricht *...den der Sohn kämmt sich im Wohnzimmer...* Auf der von uns analysierten Position haben die VTn die Namen der Agentia *Vater* und *Sohn* sowie als einziges Verb *kämmen* im Singular Präsens gehört, das damit zu beiden Agentia kongruent ist. Folglich muss die Interpretation vorwiegend auf diesen lexikalischen Informationen beruhen. Die Verteilung der Fixationen mit dem höchsten absoluten Anteil für die beiden Abbildungen LSK und nonmatch-Verb(RC) legt nahe, dass vor allem die beiden Abbildungen in Betracht gezogen und verglichen werden, die die Handlung des aktuellen Verbs darstellen (X KÄMMT SICH).

Wie wirkt sich hier die LSK auf die Interpretation aus und wie erklärt sich der Unterschied zur Kontrollbedingung? In der Kontrollbedingung wird durch das gebundene Partikelverb markiert, dass der letzte Teilsatz (der Relativsatz) abgeschlossen ist, und der kommende Teilsatz einen neuen Interpretationsaspekt einführt, in dem mit höchster Wahrscheinlichkeit das Agens des Matrixsatzes eine zentrale Rolle spielen wird. Da aber bislang nur die Handlung des Relativsatzverbs bekannt ist, wird vermutlich diese zum Wechsel der Agentia in Beziehung gesetzt. Dies könnte zum hohen Anteil der visuellen Aufmerksamkeit für die Interest Area nonmatch-Verb(RC) führen. In der LSK-Bedingung hingegen wird der Abschluss des letzten Teilsatzes nicht erkannt, denn die Sequenz wird lokal nicht als Relativsatz-Struktur verarbeitet, sondern als Hauptsatz-Struktur. Und in dieser ist nach wie vor das Agens des Relativsatzes dasjenige Agens, das die aktuelle Handlung ausführt. Diese Interpretation der Fixations-Verhältnisse und des LSK-Effekts legt nahe, dass die Fixation vor allem durch die Interpretation der Handlung gelenkt wird, die durch das aktuelle Verb beschrieben wurde. Durch die LSK wird nun manipuliert, welches Agens als Ausführender der Handlung in Betracht gezogen wird.

Eine ähnliche Erklärung deutet sich bei der Verschiebung der Fixationsanteile bei den Personalpronomen an: Im Beispiel zeigt die LSK-Abbildung die Szene SOHN KÄMMT VATER, die nonmatch-Verb(RC)-Abbildung die Szene VATER KÄMMT SOHN und die nonmatch-Verb(MC)-Abbildung die Szene SOHN ZIEHT VATER AN. Die Satzsequenz in der LSK-Bedingung lautet *...den der Sohn kämmt ihn im Wohnzimmer...* Wieder erhalten die beiden Abbildungen der Handlung des aktuellen Verbs den höchsten Anteil an Fixationen. Die Manipulation der Satzbedingungen steuert allerdings nicht, wer bevorzugt als Agens und wer als Patiens angenommen wird, denn für das aktuelle Verb ist die lokale und globale Handlung und die Verteilung der thematischen Rollen identisch. Vielmehr scheint das Erkennen der Satzgrenze und damit des Teilsatz-Abschlusses in der Kontrollbedingung zu gelingen, während die LSK-Bedingung den Versuch zur Fortführung des aktuellen Teilsatzes begünstigt. Die Lenkung der visuellen Aufmerksamkeit lässt vermuten, dass die Manipulation vor allem beeinflusst, wie früh die VTn nach einer neuen Abbildung suchen, die den neuen Interpretationsaspekt des aufkommenden Teilsatzes darstellt. Die Fortführung der Interpretation des Teilsatzes in der LSK-Bedingung führt zu einer verspäteten Reorientierung.

Merkwürdig ist allerdings der Unterschied für die nonmatch-Verb(MC)-Abbildung, die den Referenten des Relativsatzes als handelndes Agens des noch nicht gehörten Matrixsatz-Verbs zeigt. Eine spekulative Erklärung für den erhöhten Fixations-Anteil in der Kontrollbedingung könnte sein, dass durch das gebundene Partikelverb zwar die Satzgrenze erkannt und nach einer Abbildung für einen neuen Interpretationsaspekt gesucht wird, dabei aber nicht die lokale Kohärenz der Referenzstruktur durchbrochen wird. Damit bleibt der lokale Referent als potentielles Agens einer Handlung aufrechterhalten, und die Suche nach einer alternativen Szene mit neuer Handlung berücksichtigt ihn in dieser thematischen Rolle, während er in der LSK-Bedingung stärker an die Handlung des Relativsatz-Verbs gebunden ist. In dieser Interpretation kommt er allerdings nicht als Antezedens des Personalpronomens in Frage, das folglich an ein anderes Antezedens gebunden werden muss.

**Zeitlicher Verlauf** Wie kommt es im Fall der Reflexivpronomen zu einem späten und bei den Personalpronomen zu einem vergleichsweise frühen Effekt? Unsere erste Erklärung, dass die visuellen Stimuli in Experiment 3 im Vergleich zu den Stimuli aus Experiment 2 eine höhere Komplexität aufweisen, kann in dieser einfachen Form wohl nicht aufrecht erhalten bleiben, denn sie müsste für beide Bedingungen gleichermaßen gelten. Unsere an der Interpretation der Handlung orientierte Erklärung bietet zumindest für den sehr frühen Einsatz in den Bedingungen mit Personalpronomen einen Erklärungsansatz. Da der Effekt vor allem durch das Erkennen des Teilsatz-Abschlusses und seiner Interpretation gesteuert wird, und wir in Experiment 3 szenische Darstellungen der Handlungen verwenden, erfolgt der On-Set des Effekts am Relativsatz-Verb und nicht an der Anapher. Denn ab jener Position kann festgestellt werden, dass eine neue Handlung folgen muss – und da Handlungen abgebildet werden, kann auch ein Abgleich mit diesen erfolgen. Vom geschätzten On-Set des Verbs bemessen erhalten wir damit in Annäherung wieder einen On-Set des Effekts, wie er in Experiment 2 geschätzt wurde.

Allerdings müsste der Effekt auch bei den Reflexivpronomen bereits am Relativsatz-Verb einsetzen. Es ist schwer, für diesen Unterschied zwischen den Bedingungen eine einleuchtende Erklärung zu finden. Einem sehr spekulativen Ansatz folgend könnte der Unterschied auf der Verfügbarkeit starker struktureller Constraints beruhen: Das dem Relativsatz-Verb folgende Reflexivpronomen bietet einen starken strukturellen Hinweis für die Folgeinterpretation und fällt deshalb im Vergleich zum Effekt des Verbs ins Gewicht. Ein folgendes Personalpronomen hingegen bietet einen schwächeren strukturellen Hinweis auf die Folgeinterpretation. Zumal identifiziert es nur das Patiens und bietet keinen Hinweis auf das für die Handlung wichtigere Agens, weshalb es im Vergleich zum Effekt des Verbs möglicherweise nicht ins Gewicht fällt. Allerdings treten sowohl Personal- als auch Reflexivpronomen erst nach dem Relativsatz-Verb auf, weshalb sich eigentlich kein Unterschied zwischen den Pronomen zeigen sollte. Dies ist nur dann denkbar, wenn die möglichen Folgewörter stark antizipiert und die pronominalen Formen folglich schon am On-Set erkannt und unterschieden werden. Diese Annahme wird durch die unterschiedlichen visuellen Stimuli in den beiden Bedingungen des Anaphertyps bestärkt, die im einen Fall nur reflexive und im anderen Fall nur transitive Handlungen zeigen. Hier könnten anders als in Experiment 2 wiederum Interpretationseffekte durch die visuellen Stimuli induziert worden sein.

**Referenz vs. Handlung** Kann nach dieser Erklärung für die gefundenen Effekte noch beansprucht werden, dass sie sich auf die Auflösung der pronominalen Referenz auswirken? Oder anders formuliert: Macht sich der Effekt der LSK auf die Interpretation daran fest, wer über die Co-Referenz der Anapher als Agens oder Patiens der beschriebenen Handlung berücksichtigt wird? Hinsichtlich der gewichtigen Rolle des Verbs für die Interpretation wird diese Position in Frage gestellt. Denkbar wäre in diesem Zusammenhang, dass die gefundenen Effekte in keinerlei Beziehung zur Auflösung der anaphorischen Referenz stehen: Sie zeigen eher ein verspätetes Bemerken der Satzgrenze, aufgrund dessen die Anapher an den vorhergehenden Satz angebunden wird, ohne aber interpretiert zu werden. Die verspätete Feststellung der Satzgrenze führt zu einer verspäteten Suche nach einer neuen Abbildung, die zu einer möglichen Interpretation des folgenden Teilsatzes korrespondiert.

Dieser Einschätzung kann folgendes entgegengehalten werden: In beiden Bedingungen des Faktors Anapherentyp zeigt sich, dass das aktuelle Verb eher übergreifend die in Betracht gezogenen Interpretationen beeinflusst und darüber die Aufmerksamkeit auf verschiedene Abbildungen lenkt. Denn immer ist es das Verb des Relativsatzes (bzw. dessen Handlung), das bestimmt, welche Abbildungen bevorzugt betrachtet werden. Den tatsächlichen Unterschied für die Bedingungen des Satztyps und damit für die Wirkung des LSK-Effekts bestimmt aber die Auswahl des handelnden Agens. Während ein Wechsel des aktuellen Agens natürlich auch unabhängig von der Anapher erfolgen könnte, verhält es sich dennoch so, dass durch die Bedingungen ganz klar gesteuert wird, ob der lokale Referent als Antezedens berücksichtigt wird. Weiterhin besteht nach wie vor der Unterschied zwischen den beiden pronominalen Formen. Auch wenn sich dieser nur schlecht erklären lässt, so weist er dennoch darauf hin, dass die Anaphern bei der Interpretation berücksichtigt werden und auf den zeitlichen Verlauf der Interpretation einwirken.

**Problematisierung der Stimuli** Zuletzt ist noch anzumerken, dass die Wahl der Konstruktion der sprachlichen Stimuli möglicherweise wiederum neue Schwächen aufweist: Problematisch ist hier die Wahl der LSK-Durchbrechung im Kontrollsatz durch das gebundene Partikelverb, durch das eindeutig das Ende des Teilsatzes markiert wird. Dies könnte auf dieser Position möglicherweise zu Wrap-up-Effekten führen. Diese Effekte dürften in der Kontrollbedingung wesentlich stärker sein als in der LSK-Bedingung. In letzterer tauchen die Effekte vermutlich erst mit einiger Verzögerung auf, denn als gebundenes Partikelverb könnte diese Position eindeutiger als Abschluss eines Teilsatzes identifiziert werden.

Allerdings ist dieser Interpretation eine Reihe von Einwänden entgegenzuhalten: Erstens ist unklar, wie sich die Wrap-Up-Effekte auf die abhängige Variable der Visual-World-Experimente auswirken. Zweitens sollte durch die identische syntaktische Struktur eines Nebensatzes eindeutig die Abgeschlossenheit des Teilsatzes markiert sein und der morphosyntaktische Unterschied des finiten Verbs hier eine dementsprechend geringfügige Rolle spielen. Drittens: Sollte sich tatsächlich doch ein Wrap-Up-Prozess abspielen, so würde man am ehesten erwarten, dass sich durch die Effekte eine Verzögerung folgender Sakkaden und eine Verlängerung der aktuellen Fixation ergibt. Dementsprechend müssten wir eher mit einer Überlagerung der Auswirkungen des LSK-Effekts rechnen. Finden wir dennoch LSK-Effekte, so spricht dies eher für die Stärke ihrer Auswirkungen.

## 7.5. Diskussion der Befunde und Konsequenzen

Mit der Experimentalserie AnaLoCo haben wir versucht die Frage zu beantworten, ob syntaktische Subsequenzen, die isoliert im lokalen Kontext betrachtet eine kohärente Satzstruktur bilden, bei der Verarbeitung auf die interpretative Ebene der Referenzauflösung Einfluss nehmen. Wir haben für die Untersuchung Personal- und Reflexivpronomen verwendet, wobei die sprachlichen Stimuli nicht ambig sind – d.h. die Pronomen können strukturell eindeutig aufgelöst werden – und nicht in die kritische Klasse der PNPs gehören. Für die Verarbeitung von Anaphern in LSK haben wir verschiedene Hypothesen formuliert: (a) Der lokale syntaktische Kontext ist ein relevanter struktureller Einflussfaktor bei der

Anaphernauflösung, (b) strukturelle Constraints wirken nicht nur global konsistent, sondern bereits auf der lokalen Ebene, (c) die Verarbeitung auf der lokalen Ebene greift tiefer als bislang angenommen bis in die Ebene referenzieller Interpretation, (d) Anaphern werden bereits frühzeitig tief verarbeitet, deshalb ist ihre Verarbeitung generell sensitiv für LSK, (e) verschiedene Anapherntypen weisen eine unterschiedliche Sensitivität für strukturelle Einflussfaktoren auf und reagieren deshalb unterschiedlich auf den Einfluss des lokalen strukturellen Kontexts.

### 7.5.1. Zusammenfassung zentraler Ergebnisse

Die Ergebnisse aus den verschiedenen Experimentalparadigmen – Self-Paced Reading mit anschließender geschlossener Verständnisfrage und Visual-World-Paradigma mit Referenten- und Szenen-Abbildungen – beleuchten die Frage nach dem Einfluss von LSK auf die Anaphernauflösung aus verschiedenen Perspektiven. Sie gewichten dabei unterschiedliche Aspekte der Verarbeitung (visuelle Aufmerksamkeit, Lesezeit, Leseverständnis, Akzeptabilitätsbeurteilung) und der sprachlichen Stimuli (Kongruenz dependenter Einheiten, personale Referenz, Handlung des Verbs). Wir fassen kurz die zentralen Ergebnisse und Schlussfolgerungen aus den Experimenten zusammen.

Im **Self-Paced Reading-Experiment (Exp. 1)** fanden wir einerseits keinen Effekt geringerer Lesezeiten für die LSK-Bedingungen während der LSK. Wir konnten aber Tendenzen eines LSK-Effekts mit erhöhten Lesezeiten auf der Folgeposition (dem Matrixsatz-Verb) feststellen, ab der sich die lokale Subsequenz nicht kohärent fortsetzen lässt. Zudem fand sich auf der späteren Folgeposition des beginnenden Hauptsatzes ein signifikanter Effekt geringerer Lesezeiten für die LSK-Bedingungen, wobei dieser durch die unerwartete Effekt-Richtung eher schwer zu erklären ist. Da sich in den anschließenden Verständnisfragen kein Effekt für den Faktor Satztyp (im Sinne einer geringeren Quote korrekt beantworteter Fragen) fand, ist die Erklärung dieser Beschleunigung durch einen Zusammenbruch der Interpretation in dieser Bedingung mit einer anschließenden oberflächlichen und deshalb schnelleren Suche nach einem Wiederanknüpfungspunkt im Satz eher unwahrscheinlich. Ein starker Effekt zeigt sich für die Anapherntypen mit deutlich geringerer Lesezeit für Reflexivpronomen über die gesamte zweite Hälfte des Matrixsatzes. Da Reflexivpronomen strukturell eindeutig aufgelöst werden können, werden sie möglicherweise deutlich einfacher verarbeitet. Allerdings konnten wir entgegen unseren Erwartungen keine Interaktion von Anapherntyp und Satztyp finden. Die Ergebnisse decken sich weitgehend mit den Effekten für die Verständnisfragen. Hier zeigten sich für den Faktor Anapherntyp sowohl unterschiedliche Latenzen als auch eine unterschiedliche Quote korrekt beantworteter Fragen mit einem klaren Vorteil für die Bedingung Reflexivpronomen, aber keine Interaktion zwischen Satztyp und Anapherntyp. Für den Faktor Satztyp wurde nur in den Beantwortungslatenzen ein Effekt sichtbar – entsprechend unseren Erwartungen mit höherer Latenz für die LSK-Bedingung. Zusammengefasst zeigen sich Haupteffekte für Satztyp und Anapherntyp, wenngleich diese später als erwartet und in mitunter unerwarteter Ausprägung auftreten.

Grundsätzlich scheinen sich LSK auf die Verarbeitung auszuwirken und weiterhin deuten sich Verarbeitungsunterschiede für die Anapherntypen an. Allerdings konnte unsere Hypothese einer Interaktion der Faktoren nicht bestätigt werden. Eine mögliche Erklärung

für dieses Muster wäre, dass sich LSK nur auf die frühe Verarbeitung im Sinne eines initialen Filters auswirken, der auf beide Anapherntypen in gleicher Weise wirkt. Allerdings könnte dieser auch nur einen beschränkten Ausschnitt der Auswirkungen zeigen, der durch die Eigenschaften der verwendeten Methode Self-Paced Reading begünstigt bzw. hervorgehoben wurde, da sich hier Reanalyse-Mechanismen nicht klar zuordnen lassen und sich durch die künstliche Lesesituation vor allem initiale Verarbeitungseffekte auf die Lesezeitmaße auswirken. Die Frage des Verlaufs der referenziellen Auflösung (d.h. wer wird an welcher Stelle als Antezedens berücksichtigt) lässt sich mit dieser Methode jedenfalls nicht beantworten.

Die erste **Visual-World-Studie (Exp. 2)** erweitert unsere Befunde aus der Lesezeitmessung. Hier zeigt sich, dass es sich bei den gefundenen Effekten nicht nur um syntaktische Interferenzen, sondern um Effekte handelt, die die Interpretation und darüber wiederum die visuelle Aufmerksamkeit beeinflussen. Für das Reflexivpronomen zeigen sich in der LSK-Bedingung erhöhte Fixationsanteile auf dem Relativsatz-Subjekt bei gleichzeitiger Verminderung der Fixationen auf dem unbeteiligten Referenten. Beim Personalpronomen zeigen sich keine Effekte auf den Antezedenzen, dafür aber eine signifikante Erhöhung der Fixationsanteile auf dem unbeteiligten Referenten. Das könnte dadurch erklärt werden, dass bei (lokal) ambigem strukturellem Kontext der Kreis der berücksichtigten Kandidaten für eine Koreferenz-Beziehung ausgeweitet und die Auflösung stärker durch nicht-strukturelle Faktoren geleitet wird.

Auch mit der Eyetracking-Methode wurden also Effekte der LSK gefunden. Die beschränkten Effekte im Fall der Personalpronomen im Vergleich zu den Reflexivpronomen könnte einerseits durch die von uns unterstellte unterschiedliche Sensitivität der Anapherntypen für strukturelle Einflüsse erklärt werden. Andererseits wird die Erklärbarkeit der Befunde dadurch eingeschränkt, dass weiterhin Unklarheit herrscht, durch welche Aspekte des linguistischen Inputs die visuelle Aufmerksamkeit gesteuert wird – zum Beispiel eher durch Referenz, durch thematische Rollen oder durch die mit dem Verb beschriebene Handlung. Die Interpretation der Befunde wird zusätzlich durch die Ergebnisse der Rating-Studie eingeschränkt. In dieser finden sich Hinweise darauf, dass die Manipulation der auditiven Stimuli – die Eliminierung prododischer Marker über eine Satzgrenze hinweg – zu einer schlechteren Beurteilung der Natürlichkeit und Verständlichkeit der Stimuli führt. Diese Manipulation könnte die Berücksichtigung und Verarbeitung von LSK in unrealistischer Weise begünstigen.

Die zweite **Visual-World-Studie (Exp. 3)** führt eine Verbesserung der sprachlichen Stimuli und eine Variation der visuellen Stimuli ein. Hier zeigen sich nun sowohl für die Reflexivpronomen als auch für die Personalpronomen Effekte erhöhter Fixationsanteile in der LSK-Bedingung. Die Fixationen finden sich hier vermehrt auf den Szenen, die die LSK abbilden. Im Fall der Personalpronomen, für die die globale und die lokale Interpretation deckungsgleich ist, äußert sich dies in einer verlängerten gerichteten Aufmerksamkeit auf die Abbildung der Handlung des Relativsatz-Verbs. Die erhöhten Fixationsanteile auf den zur LSK korrespondierenden Abbildungen gehen in erster Linie auf Kosten der Abbildungen, die weder lokal noch global mit den Verteilungen der thematischen Rollen im Satz korrespondieren. Beim Reflexivpronomen ist dies die Szene, die die Handlung des Relativsatz-Verbs mit falschem Agens abbildet. Dies legt eine starke Steuerung der visuellen Aufmerksamkeit durch die Interpretation des Verbs nahe, wobei die LSK aber

beeinflusst, welcher Referent als Agens der hier beschriebenen Handlung berücksichtigt wird. Beim Personalpronomen wird der Effekt auf die Aufmerksamkeitsverteilung aufgrund der übereinstimmenden lokalen und globalen Interpretation vor allem in Abhängigkeit davon geleitet, wann der Abschluss des Teilsatzes erkannt und in der Folge eine neue Szene für den aufkommenden Teilsatz gesucht wird. Die verlängerte Aufmerksamkeit auf der lokalen Szene geht auf Kosten der Szene für das Matrixsatz-Verb mit falschem Agens. Sie kann möglicherweise durch die stärkere Bindung des Relativsatz-Subjekts an die Handlung des Relativsatz-Verbs bei der initialen Interpretation des aufkommenden Teilsatzes erklärt werden.

Eher unerwartet ist die zeitliche Ordnung der Effekte, die genau entgegengesetzt zu unseren Erwartungen verläuft: Während der sehr frühe Einsatz des Effekts im Fall der Personalpronomen gut zu unserer Erklärung passt, dass die visuelle Aufmerksamkeit stark durch die Interpretation des Verbs geprägt wird, wird der eher späte Effekt beim Reflexivpronomen weniger gut durch diesen Ansatz abgedeckt. Eine mögliche Erklärung könnte aus der Verbindung eines sehr frühen Effekts der Anaphernverarbeitung durch visuell induzierte Antizipation des Anapherntyps mit der Prägung der visuellen Aufmerksamkeit durch die Verbhandlung entwickelt werden, die sich – möglicherweise ebenfalls durch die visuellen Stimuli induziert – unterschiedlich für die beiden Anapherntypen auswirkt. Die LSK bedingt bei den Personalpronomen eine Reduzierung der Fixationsanteile auf der Szene für das Matrixsatz-Verb und bei den Reflexivpronomen eine Reduzierung auf der Szene für das Relativsatz-Verb. Es scheint, als würde das Reflexivpronomen (oder aber die visuellen Stimuli) hinsichtlich des darauf bezogenen Verbs eine stärker rückwärtsgerichtete Interpretation nahelegen, was zu einem späteren Einsatz der Suche nach der Interpretation eines aufkommenden nächsten Teilsatzes führt. Insgesamt weisen die Befunde aus Experiment 3 auf eine starke Gewichtung der Interpretation des Verbs hin – hinsichtlich der Verteilung der visuellen Aufmerksamkeit sowie hinsichtlich des Einflusses der LSK. Allerdings zeigt sich in den Details, dass auch die Referenzauflösung mit Aufmerksamkeit und LSK interagiert, was sich im unterschiedlichen zeitlichen Verlauf sowie in den unterschiedlichen betroffenen IAs (Abbildungen der Handlung des Matrixsatz-Verbs vs. des Relativsatz-Verbs) äußert.

### 7.5.2. Bezug zu Theorien für LSK und Anaphernaauflösung

Die Ergebnisse der Studie haben in verschiedener Hinsicht Konsequenzen für Theorien der LSK sowie der Anaphernverarbeitung und die dort getroffenen Annahmen. Wir diskutieren diese Bezüge in den folgenden Abschnitten.

**Lokale syntaktische Kohärenzen** Den Ergebnissen folgend nehmen wir an, dass LSK-Effekte nicht nur ein Störfaktor der globalen Interpretation durch lokale Subsequenzen sind. Vielmehr scheint die Verarbeitung in lokalen Zeit- oder Sequenzfenstern eine inhärente Eigenschaft der zugrundeliegenden Verarbeitungsmechanismen zu sein, die auch in der allgemeinen Verarbeitung wirkt. Dabei wirkt sie sich aber eben nur dann negativ auf die globale Interpretation aus, wenn die Stimuli lokale Subsequenzen enthalten, die in der lokalen Verarbeitung der globalen Interpretation widersprechen. Zudem scheint die Verarbeitung in lokalen Fenstern eine eher generelle kognitive Eigenschaft zu sein, die sich

auf verschiedene Interpretationsaspekte und beteiligte Mechanismen auswirkt. Deshalb betrifft sie nicht nur die oberflächliche syntaktische Anbindungsebene, sondern auch die Ebenen der semantischen Verarbeitung, auf der Handlung, Referenz und thematische Rollen interpretiert werden. Für all diese verschiedenen Aspekte der Verarbeitung wird dabei in einem frühen Stadium die lokale Verarbeitung vorübergehend isoliert berücksichtigt.

Aus dieser Perspektive können wir die Ergebnisse kritisch in Beziehung zu verschiedenen Theorien und Annahmen über LSK-Effekten setzen, die wir in Kapitel 5.2.4 besprochen haben:

- Beschaffenheit des verarbeiteten Kontexts: Können die Effekte der Berücksichtigung falscher Antezedenzen alleine aus der Interferenzen intervenierender Agentia erklärt werden, wie von van Dyke (2007) nahegelegt wird? Wir können eine allgemeine Auswirkung intervenierender Agentia mit unseren Daten zwar nicht ausschließen, denn hierfür haben wir keine Kontrollbedingung. Allerdings zeigen die Befunde durch unsere Kontrollbedingungen, dass die syntaktische Kohärenz der Subsequenz Effekte erzeugt, die über eine bloße Anwesenheit anderer Agentia hinausgeht. Denn das intervenierende Agens ändert sich in unseren Bedingungen nicht.
- Lexikalisch-statistische vs. interpretative Effekte: Sind die Effekte durch rein lexikalische Merkmale erklärbar, etwa durch bottom-up Statistiken oder durch den Abgleich von sequenzbezogenen Bottom-up- und Top-down-Statistiken (R. L. Lewis et al., 2005; Gibson, 2006; Bicknell & Levy, 2009; Hale, 2011)? Beim Gegenstand unserer Studie stoßen Erklärungen auf dieser Basis an ihre Grenzen, denn die lexikalische Wahrscheinlichkeit für Reflexiv- und Personalpronomen hat nichts mit der koreferenziellen Wahrscheinlichkeit der möglichen Referenten zu tun, bzw. aus den verschiedenen lexikalischen Wahrscheinlichkeiten der pronominalen Formen in unseren Satzbedingungen lässt sich keine Wahrscheinlichkeit für die referenzielle Bindung ableiten. Diese Ansätze bieten also keine umfassende Erklärung für unsere Befunde, für die wir demnach eine tiefere interpretative Verarbeitung annehmen müssen.
- Visuelle Induzierung: In Beziehung zu den früheren Studien zu LSK mit dem Visual-World Paradigma (Konieczny & Müller, 2007; Konieczny, Müller, Hachmann et al., 2009) und ihrer Interpretierbarkeit bzw. ihrer Kontrolle gegenüber Alternativerklärungen wird unser erstes Experiment für die Erforschung dieses Gegenstands wichtig: Mit der kontextfreien Abbildung von Einzelreferenten, die keinen Rückschluss darüber erlauben, in welche Handlung sie eingebunden sind, sind die dargebotenen Stimuli situational neutral.<sup>54</sup> Durch diese Neutralität der Abbildungen gegenüber der Interpretation des Satzes ist eine Induzierung der Interpretation der LSK-Lesart ausgeschlossen. Damit zeigen unsere Befunde, dass LSK Effekte sind, die unabhängig von der Natur der visuellen Stimuli erfolgen. Auf ähnliche Weise stützen auch die Ergebnisse von Müller et al. (2009), die anhand von Blickbewegungsaufzeichnung

---

<sup>54</sup>Wie in der Diskussion in Kapitel 5.2.3 erläutert, besteht die Möglichkeit, dass in den vorausgegangenen Studien nur deshalb LSK-Effekte gefunden wurden, weil die visuellen Stimuli durch Abbildung der in der LSK beschriebenen Situation eine entsprechende Interpretation der Sequenz nahe legen. Eine solche Induzierung der Berücksichtigung der LSK-Lesart ist mit den in diesem Experiment gewählten visuellen Stimuli aber nicht möglich, denn die kontextfrei abgebildeten Referenten legen keine Lesart für ihre thematische Rolle oder eine bestimmte Handlung nahe.

während des Lesens mit rein textuellen Stimuli erhoben wurden, diesen Befund kontextunabhängiger LSK-Effekte.

- Tiefe der Verarbeitung: Wir gehen davon aus, dass LSK-Effekte weit mehr sind als die von Tabor et al. (2004) angenommenen Interferenzen an der syntaktischen Oberfläche. Verschiedene Befunde zur visuellen Aufmerksamkeit – aus der hier dargestellten und aus verschiedenen weiteren Studien (u.a. Konieczny & Müller, 2007; Konieczny, Müller, Hachmann et al., 2009; Weldle, Konieczny, Wolfer et al., 2010) – legen nahe, dass sich LSK auf eine Verarbeitungstiefe bis zur Ebene der thematischen, referenziellen und diskursiven Interpretation hin auswirken.
- Shallow- oder Good-Enough-Parsing: Es besteht die Möglichkeit, dass die Interpretation des lokalen Kontexts entgegen der globalen Lesart durch die hohe Komplexität der sprachlichen Stimuli zustande kommen, die zu einer Verwerfung syntaktischer Informationen und einer oberflächlichen Interpretation der Sätze führt (etwa im Sinne von Ferreira et al., 2002; Ferreira & Patson, 2007). Der mittelmäßigen Rate der korrekt beantworteten Fragen in der Lesezeitstudie (durchschnittlich wurden über 70% der Fragen korrekt beantwortet) folgend könnte diese Erklärung in Betracht kommen. Aber auch unabhängig davon ist der Ansatz des Good-Enough-Parsing ein interessanter Kandidat für die Erklärung der Effekte.

**Anaphernauflösung** LSK sind auch allgemein ein interessantes Werkzeug für den Einsatz in der psycholinguistischen Forschung. Unserer Einschätzung nach ist das Feld der damit untersuchbaren Gegenstände weitaus größer als bislang umgesetzt. Zum Beispiel stehen sich in der Psycholinguistik nach wie vor Ansätze gegenüber, die eine eher seriell-modulare oder aber eine parallel-interaktive Verarbeitung von Sätzen annehmen. Das Phänomen LSK könnte hier weitere interessante Möglichkeiten bieten, um den zeitlichen Verlauf der Verarbeitung verschiedener linguistischer Ebenen zu untersuchen. LSK-Effekte wirken vor allem im frühesten Zeitraum der Online-Verarbeitung. Entsprechende Effekte können in empirischen Studien möglicherweise als Indikator dafür dienen, wie schnell welche Aspekte des linguistischen Inputs in welcher Tiefe verarbeitet werden. Die Befunde unserer Studie weisen auf eine recht tiefe Interpretation der LSK hin. Dies deutet umgekehrt darauf hin, dass Anaphern bereits frühzeitig einer initialen referenziellen Verarbeitung unterzogen werden.

Unseren Befunden nach zu urteilen, scheint diese frühe Interpretation allerdings nicht endgültig auswählend, sondern eher vorläufig einschränkend zu sein – eventuell im Sinne eines initialen Filters zur Auswahl möglicher Antezedenzen. LSK scheinen sich aus dieser Perspektive auf die Referenzzuweisung eher auf die initiale Verarbeitung auszuwirken und die Interpretation nur kurzfristig und vorübergehend zu beeinflussen. Sie beeinflussen nur die initiale ausschließende Verarbeitung, die die globale Konsistenz der Interpretation noch nicht berücksichtigt. Während der initial einschränkenden Verarbeitung führen LSK dazu, dass durch die lokale Ambiguität weniger Referenten als Antezedens ausgeschlossen werden können. Für den initialen Filter wären nach unseren Daten andere Eigenschaften anzunehmen, als sie in der Early-Filter-Hypothese (Nicol & Swinney, 1989) oder im zweistufigen Ansatz von Sturt (2003) mit früher struktureller Auswahl vermutet werden: (a) Er berücksichtigt nicht ausschließlich strukturelle Constraints – stattdessen werden

auch semantische und kontextuelle Informationen zum Beispiel aus der visuellen Situation oder dem Diskurs berücksichtigt (vgl. auch Runner et al., 2006), (b) er arbeitet nicht ausschließlich im globalen, sondern auch im lokalen Kontext konsistent, (c) dabei können Referenten in die Auswahl treten, die zu einer Verletzung globaler Bindungs-Constraints führen würden.

Wir haben in Kapitel 6.6 die Eigenschaft von SRNs diskutiert, die lokalen Online-Effekte bei der Anaphernauflösung zu erfassen, nicht aber die abschließende Auswahl des Antezedens oder seine Eingliederung in die globale Interpretation. Gerade damit scheinen sie sich besonders gut für die Modellierung der in den Experimenten gezeigten Effekte zu eignen. Denn sie erfassen die Eigenschaften eines initialen Filters, der für die frühen Effekte bei der Anaphernauflösung verantwortlich ist, und der zu den gezeigten LSK-Effekten führt. Während für die Modellierung weiterer und vor allem später Effekte anaphorischer Verarbeitung Erweiterungen des Modells benötigt werden, eignet sich das untersuchte Modell AnaSRN gut für Simulationen initialer Auswahlprozesse der frühen Verarbeitung.

In der initialen Verarbeitung, die wir mit unserem experimentellen Vorgehen beobachten konnten, zeigen die Ergebnisse des Weiteren durch LSK verursachte Verletzungen der globalen Konsistenz der referenziellen Bindung. Auf solche Effekte wurde in verschiedenen früheren Studien hingewiesen (u.a. Runner et al., 2003, 2006). Entscheidend ist, dass unsere Befunde nicht auf dem sowieso als problematisch identifizierten Gegenstand der PNPs beruhen. Die Ergebnisse stellen die Annahme globaler Konsistenz von Bindungs-Constraints mit einer neuen Satzstruktur in Frage und erweitern damit das Feld der für diese Annahme problematischen Strukturen. Zugleich decken sich unsere Ergebnisse aber im lokalen Kontext mit den Annahmen einer strukturell geprägten Bestimmung des Antezedens. Dies legt nahe, dass die Annahmen der Bindungstheorie nicht an sich falsch sind, für die Online-Verarbeitung aber eine Sensitivität der beteiligten Prozesse für den lokalen Kontext bzw. eine lokal beschränktes Fenster bei der initialen Verarbeitung angenommen werden muss.

In diesem Zusammenhang wäre auch die in anderen Studien gezeigte unterschiedliche Sensitivität für verschiedene Anapherntypen (vgl. u.a. Kaiser et al., 2009) zu erwähnen. Zwar wurden in unserer Studie nur syntaktische Bedingungen variiert, doch auch für diese zeigen sich unterschiedliche Auswirkungen auf Reflexiv- und Personalpronomen. Für diese pronominalen Formen wirken sich die strukturellen Constraints sehr unterschiedlich aus, im einen Fall präskriptiv, im anderen restriktiv. Um eine abschließende Interpretation bzw. Auswahl des Antezedens zu erhalten, erfordern die beiden Formen in unterschiedlichem Ausmaß den Einbezug weiterer Informationsquellen, etwa morphosyntaktischer, semantischer oder pragmatischer Natur. Auf diesen Unterschied des Anteils struktureller Informationen an der Interpretation und seinen Einfluss auf die Verarbeitung weisen auch die Simulationsergebnisse im Projekt AnaSRN hin, in dem sich langfristig eine geringere Verbesserung der Zuweisungspräzision im Fall der Personalpronomen zeigt. Diese unterschiedliche Erfordernis weiterer einschränkender Informationen bietet einen Ansatz für die Untersuchung und Erklärung von Details zur Hypothese formenspezifischer multipler Constraints (Brown-Schmidt et al., 2005; Kaiser et al., 2009). Möglicherweise kann die

konnektionistische Modellierung hier helfen, den formenspezifischen Anteil verschiedener Constraints näher zu bestimmen.<sup>55</sup>

**Bezug zur Modellierung AnaSRN** Wie sind unsere konnektionistischen Modellierungen auf die Befunde der empirischen Studien hin zu beurteilen? Die Effekte, die wir für die Wortvorhersage-Aufgaben beobachten konnten, wurden bereits in früheren konnektionistischen Studien demonstriert und erfolgreich mit empirischen Befunden der Psycholinguistik in Verbindung gebracht (vgl. u.a. Weckerly & Elman, 1992; Tabor et al., 1997; Christiansen & Chater, 1999b; MacDonald & Christiansen, 2002; Konieczny, Ruh & Müller, 2009; Wells et al., 2009; Konieczny & Müller, 2010). Mit den hier vorgestellten Studien lassen sich erweiternd erste Bezüge zwischen Ergebnissen empirischer Forschung und konnektionistischer Modellierung für den Gegenstand der Anaphernauflösung herstellen. Über die Simulationen im Projekt AnaSRN wurden LSK-Effekte auch für die Anaphernauflösung vorhergesagt, die zu einer Verletzung der Annahme globaler Konsistenz einhergehen. Zudem wurde ein Unterschied zwischen Reflexiv- und Personalpronomen erwartet, und fehlende Satzstruktur-Marker sollten sich verstärkend auf den LSK-Effekt auswirken. Für all diese modellierungsgeleiteten Annahmen haben wir korrespondierende Befunde in unseren Daten gefunden. Damit unterstützen die Ergebnisse die Berechtigung einer konnektionistisch ausgerichteten Psycholinguistik.

---

<sup>55</sup>Eine alternative Erklärung für die unterschiedlichen Effekte könnte sein, dass in den Bedingungen mit Personalpronomen Alternativkonstruktionen erwartet werden: Für den Verweis auf einen direkt zuvor erschienenen Referenten kann man auch Demonstrativpronomen verwenden, die dementsprechend bei der Satzverarbeitung erwartet werden. Die Existenz solcher Alternativen könnte die Ursache für eine erschwerte Verarbeitung der Personalpronomen in der gegebenen syntaktischen Konstruktion sein (vgl. Baumann, Konieczny & Hemforth, 2011; Hemforth, Scheepers, Konieczny, Pynte & Colonna, 2009; MacWhinney, 2008; Gibson, 2006). Dies könnte auch den Effekt erhöhter Lesezeiten auf den Personalpronomen im Experiment AnaLoCo 1 erklären.



## **Teil III.**

# **Konnektionismus, Embodiment und Sprachverstehen**



## 8. Embodied Cognition und Sprachverarbeitung

Etwa zeitgleich zum (Neo-)Konnektionismus hat sich eine weitere theoretische Konzeption der Kognition entwickelt (bzw. weiterentwickelt), die den klassischen symbolbasierten Theorien der Kognition kritisch gegenübersteht: Die Embodied Cognition. Überträgt man die Konzepte dieses Ansatzes auf den Gegenstand Sprache, zeigen sich viele Facetten einer völlig neuen Perspektive auf Sprachverarbeitung und Sprachverstehen mit spannenden Konsequenzen. So ist es nicht verwunderlich, dass die Forschungsaktivitäten auf diesem Gebiet in den letzten Jahren deutlich zugenommen haben und sich Symposien auf größeren Tagungen und mittlerweile sogar eigenständige Konferenzen zum Thema etablieren.<sup>1</sup> Für uns ist es von besonderem Interesse, dass sich viele der hier entwickelten Ansätze konzeptionell mit Ansätzen des Konnektionismus decken oder zumindest mit diesen überlappen. Vielfach werden konnektionistische Modelle zu Konzepten der Embodied Cognition in Beziehung gesetzt, speziell zum Thema Sprachverarbeitung. Doch inwiefern sind Konnektionismus und Embodied Cognition komplementäre Rahmentheorien bzw. Forschungsprogramme? Inwiefern kann man die Embodied Cognition überhaupt mit computergestützten Modellen erfassen? Und inwiefern bietet umgekehrt die Embodied Cognition eine Grundlage für eine genuin konnektionistische Konzeption von Bedeutung und Sprachverstehen? In den folgenden beiden Kapiteln versuchen wir, diese Fragen zu erörtern und gemeinsame Grundlagen von Embodied Cognition und Konnektionismus zu identifizieren.

Das erste Kapitel beschäftigt sich mit der Perspektive, die die Embodied Cognition auf Sprache und Sprachverarbeitung wirft. Wir beginnen mit einem kurzen Abriss zur Embodied Cognition: Nach einigen Grundlagen und einer kurzen Erläuterung zentraler philosophischer Probleme, an denen diese Position ansetzt, wenden wir uns der Frage der Konzeption kognitiver Prozesse und Repräsentationen sowie ihrer Abgrenzung von klassischen Theorien der Kognition zu. Der Schwerpunkt wird dabei auf den Konzeptionen von Glenberg (1997) und Barsalou (1999) liegen. Ziel ist die Identifikation zentraler Eigenschaften – darunter Analogizität, Multimodalität, Aufhebung der Trennung zwischen Kognition, Perzeption und Motorik sowie die Konzeption von Denkprozessen als Simulationsprozesse. In einem zweiten Abschnitt werden wir uns der Perspektive auf Sprache und Sprachverstehen nähern: Sprache wird aus Sicht der Embodied Cognition als Konstruktionsanweisung perzeptueller und motorischer Simulationen betrachtet – beim Verstehen simulieren wir Derivate früherer Erfahrungen. Das Verstehen eines Satzes entspricht dem erfolgreichen Aufbau eines mentalen Modells der Referenzsituation aus Konglomeraten

---

<sup>1</sup>Zu nennen wären zum Beispiel die Embodied Sentence Processing (ESP) 2007 in Saarbrücken, die Embodied and Situated Language Processing (ESLP) 2009 in Rotterdam, sowie in der jüngeren Vergangenheit die Etablierung eigenständiger Symposien auf größeren Konferenzen (u.a. AMLaP, CogSci und TeaP).

partieller perzeptueller und motorischer Erfahrungen. Dabei werden wir Ansätze zu einer Rahmentheorie des Embodied Language Processing (Barsalou, 1999; Glenberg & Robertson, 2000; Zwaan, 2004; Bergen & Chang, 2003; MacWhinney, 2008) diskutieren. Anschließend folgt eine kurze Zusammenfassung der wichtigsten empirischen Ergebnissen, die im Rahmen der Erforschung des Embodied Language Processing gezeigt wurden. In den psycholinguistischen Untersuchungen wurden Bezüge zwischen Sprachverarbeitung und der Bearbeitung perzeptuell-sensorischer sowie motorischer Aufgaben hergestellt. Daneben wurden auch Befunde der Gedächtnispsychologie adaptiert, die Effekten mentaler Situationsmodelle zugeschrieben wurden. Funktional-neuroanatomische Studien ergänzen diese Befunde über Zusammenhänge von Korrelaten der Hirnaktivität für sprachliche, perzeptuelle und motorische Ereignisse. Beide methodischen Zugangswege weisen auf die Simulation von perzeptuellen und motorischen Erfahrungen als essentielle Bestandteile des Sprachverstehens hin. In einem letzten Abschnitt stellen wir aktuelle Studien vor, die sich mit der Perspektive bei der mentalen Simulation während des Sprachverstehens auseinandersetzen.

### 8.1. Embodied Cognition

Ansätze der Embodied Cognition versuchen, kognitive Prozesse in Abhängigkeit des physischen Systems zu verstehen, aus dem sie hervorgehen. Man kann diesen Ansatz somit als klare Absage an den Funktionalismus verstehen – Kognition kann nicht implementationsunabhängig auf formal-abstrakter Ebene verstanden werden. Die Mechanismen und Eigenschaften der Kognition resultieren aus den Eigenschaften der Implementierung des Systems und aus der Interaktion dieses Systems mit der Umwelt. Die Herausbildung der Eigenschaften aus der Interaktion mit der Umwelt stellt die Embodiment-Ansätze in Zusammenhang mit Emergenz-Ansätzen, mit denen wir uns später noch beschäftigen werden.

Unter das Forschungsprogramm der Embodied Cognition werden verschiedene Aspekte und Gegenstände kognitiver Verarbeitung und sehr unterschiedliche Konzeptionen von Kognition subsumiert. Die Ansätze reichen von der bloßen Berücksichtigung perzeptioneller Eigenschaften des Körpers für die Verarbeitung über die Erklärung kognitiver Prozesse über körpergebundene Eigenschaften (Embodied Cognition) oder die Erklärung kognitiver Prozesse aus ihrer situativen Einbettung und der Interaktion mit der Umwelt (Situated Cognition) bis zur holistischen Sichtweise von Nervensystem, Peripherie und Umwelt als ein einziges umfassendes kognitives System (Distributed Cognition) ohne Grenze zwischen Individuum und Umwelt (vgl. Shapiro, 2007; Rohrer, 2005; Wilson, 2002; MacWhinney, 1999b). In allen Ausprägungen wird auf verschiedene Weise versucht, (a) Kognition in Abhängigkeit ihrer Implementierungsgrundlage zu erklären – dem organischen Körper mit seinen spezifischen Eigenschaften in Perzeption und Motorik, und (b) Kognition in Abhängigkeit ihrer Einbettung und Interaktion mit der Umwelt zu erklären – die ebenfalls spezifische Eigenschaften aufweist (u.a. Räumlichkeit, Zeitlichkeit, Gravitation, Temperatur, spezifische Gefahren, soziales System) und umgekehrt spezifische Eigenschaften vom Individuum einfordert (u.a. Handlungsfähigkeit in Echtzeit, Orientierung im Raum, Überlebensfähigkeit hinsichtlich Ernährung, Sicherheit, Fortpflanzung).

In den Ansätzen (u.a. Embodied, Emergent, Situated, Distributed Cognition) wird dabei eine jeweils unterschiedliche Perspektive mit unterschiedlichem Fokus auf die Eingebettetheit von Kognition in einen mit seiner Umwelt interagierenden Körper gewählt. Ausgangspunkt und Zentrum der wissenschaftlichen Theoriebildung bleibt aber die Eingebettetheit. Damit unterscheiden sich diese Ansätze von klassischen Kognitionstheorien, die die kognitive Verarbeitung losgelöst von dieser Einbettung betrachten und Kognition als von Umwelt und Körper unabhängiges System abstrakter Denkopoperationen untersuchen. In den Worten von (Wilson, 2002, S. 625): „Proponents of embodied cognition take as their theoretical starting point not a mind working on abstract problems, but a body that requires a mind to make it function.“ Daraus erklären sich die Eigenschaften unseres kognitiven Systems: Struktur und Eigenschaften des Körpers schaffen Constraints für die Interaktion mit der Umwelt, die durch das kognitive System innerhalb dieser Constraints hinreichend gut gesteuert werden muss. Die Umwelt schafft die Notwendigkeit unseres Handelns und bietet bestimmte Interaktionsmöglichkeiten und Aufgaben, die das kognitive System erfüllen muss. Unsere kognitiven Eigenschaften haben sich also in Abhängigkeit dieser Constraints und dieser lebensnahen Aufgaben entwickelt. Wollen wir abstraktere Denkprozesse erklären, müssen wir sie aus der Perspektive dieser Entstehungsgeschichte begreifen.

Die Eingebettetheit des kognitiven Systems bestimmt nicht nur dessen Eigenschaften bei der Interaktion im gegenwärtigen zeitlich-räumlichen Kontext, vielmehr sind kognitive Prozesse und mentale Repräsentationen dauerhaft verfügbare Korrelate der Interaktion. Aus dieser Perspektive eröffnet sich auch die Untersuchung von Offline-Prozessen der Kognition. Dieselben Repräsentationen und Prozesse, die unserer Online-Kognition zugrunde liegen, werden durch das kognitive System auch für die Offline-Verarbeitung verfügbar gemacht. Sensomotorische Ressourcen werden verdeckt zur Bewältigung abstrakter Aufgaben eingesetzt, und die Kategorisierung und Konzeptualisierung der Welt entspricht der mentalen Konstruktion von Repräsentationen aus Korrelaten gesammelter früherer Erfahrungen unserer Wahrnehmung und Interaktion mit der Umwelt. Die Embodied Cognition erklärt also nicht nur die Reaktion auf gegenwärtige Ereignisse und die Adaption der gegenwärtigen Handlung, sondern auch Denkprozesse, die sich mit zeitlich-räumlich abwesenden Gegenständen und Ereignissen befassen. Das kognitive System kann also entkoppelt von physischem Input und Output agieren, und ermöglicht unter anderem Leistungen wie abstraktes Denken und Planen, Erinnerung, Kreativität und Antizipation. Auf diesem Aspekt der Embodiment-Theorie – Embodied Cognition als Grundlage zur Erklärung der Eigenschaften und Mechanismen unseres Offline-Denkens sowie unserer allgemeinen Kognition – ruht der Fokus dieser Arbeit. Diesen Ansatz werden wir im weiteren als die starke Embodied-Cognition-These bezeichnen.

### 8.1.1. Theoretische Grundlagen

Die Ansätze der Embodied Cognition stehen in Beziehung zu einer Reihe theoretischer Debatten: Sie nehmen eine distinktive Position zur Natur mentaler Zustände (Leib-Seele-Problem) ein, die dem Funktionalismus als abstrakter funktionaler Erklärung mentaler Prozesse kritisch gegenübersteht und dem Trend der Rückkehr zur Identitätstheorie folgt. Eng mit dieser Frage verknüpft ist die Problematisierung der Beziehung zwischen Kogniti-

on und Umwelt und ihrer interpretativen Verknüpfung (Grounding-Transduction-Problem), in der die Embodied Cognition in der phänomenologischen Tradition argumentiert. Hinsichtlich der Frage, was Bedeutung ist, wird die Embodied Cognition in Beziehung zum Pragmatismus gesetzt. Hinsichtlich der Repräsentation bedeutungshafter Zustände und der darüber geführten Imagery-Debatte werden mitunter weitere Vorgänger (etwa empiristische Positionen) genannt. Wir greifen kurz einige dieser Debatten auf.<sup>2</sup>

Das Leib-Seele-Problem beschäftigt sich mit dem Zusammenhang zwischen körperlich-materiellen und geistig-immateriellen Prozessen und Zuständen. Die Systeme Leib (physischer materieller Körper) und Seele (mentaler immaterieller Geist) werden als in sich geschlossene Systeme mit eigenen Gesetzmäßigkeiten wahrgenommen. Wie kann dann zwischen diesen eine kausale Beziehung entstehen, d.h. wie können diese aufeinander einwirken und ein Gedanke eine materielle Wirkung erzeugen (z.B. eine mental getroffene Handlungsentscheidung ihre physische Umsetzung mit dem Körper)? Wie sind Denkprozesse und Prozesse des denkenden Körpers – in erster Linie des zentralen Nervensystems – verbunden? Dieses Problem steht im Kern der Philosophie des Geistes und ist eine zentrale philosophische Grundfrage der Kognitionswissenschaft. Eine grundlegende Position wurde mit der Identitätstheorie vorgeschlagen: In ihr wird angenommen, dass physische und mentale Zustände identisch sind – ein mentaler Zustand einer Person (etwa ein Gedanke, ein Wunsch, eine Erinnerung oder auch Schmerz) ist faktisch ein neuronaler Aktivierungszustand im Gehirn dieser Person. Aufgrund verschiedener Probleme dieses Ansatzes, etwa der Möglichkeit multipler Realisierungsformen mentaler Zustände, wurde als Alternative zur Identitätstheorie der Funktionalismus entwickelt. In diesem wird zwar weiterhin die Position vertreten, dass mentale Zustände auf physische reduziert werden können. Die konkrete materielle Realisierung ist dabei aber unerheblich: Verschiedene physische Zustände können die gleichen funktionalen Zustände erfüllen, d.h. jeder distinkte physische Zustand lässt sich einer distinkten Funktion zuordnen. Mentale Zustände sind identisch zu diesen funktionalen Zuständen. Der Funktionalismus postuliert also die Rechtfertigung einer abstrakten, rein funktionalen Betrachtungsweise mentaler Prozesse (vgl. Putnam, 1975; Fodor, 1976). Damit ist jedes System mit abstrakt beschreibbaren, funktionalen Zuständen ein kognitives System, d.h. die Eigenschaften des kognitiven Systems sind unabhängig von dessen materieller Implementierung. Während der Funktionalismus kognitionswissenschaftlichen Theorien lange Zeit beinahe axiomatisch zugrunde lag, wurden in der Zwischenzeit vermehrt Alternativen dazu formuliert. So wird etwa eine Rückkehr zur Identitätstheorie gefordert, oder aber es wird – bei einer Bewahrung der Annahme der Begründung mentaler aus physischen Zuständen – die Reduzierbarkeit auf physische Zustände abgelehnt (so etwa in Emergenztheorien). Auch in den Embodied-Cognition-Theorien lässt sich die funktionalistische Position in der Regel nicht aufrecht erhalten. Die Annahme, dass der Körper mit seinen Eigenschaften, Möglichkeiten und Beschränkungen eine zentrale Rolle für das kognitive System spielt und fundamental die Qualität und

---

<sup>2</sup>In der neueren Embodied-Cognition-Forschung finden sich viele weitere Verweise auf philosophische Vorgänger und Traditionen, auf die wir hier nicht näher eingehen werden. Vielfach wird zum Beispiel auf die Phänomenologie (v.a. Maurice Merleau-Ponty) als Ursprung des Ansatzes der Embodied Cognition verwiesen, in der Denken und Erkenntnis als in Körper und Wahrnehmung verwurzelt betrachtet wird (vgl. z.B. Lakoff & Johnson, 1999). Bezüge der Embodied Cognition zum Pragmatismus William James' werden von Rohrer (2001) diskutiert. Eine Übersicht über zugrundeliegende philosophische Positionen bieten Chrisley und Ziemke (2003), M. Anderson (2003) oder Rohrer (2007).

Eigenschaften unserer mentalen Zustände bestimmt, erzwingt die Rückkehr zur Annahme der Implementierungsabhängigkeit des kognitiven Systems: Mentale Prozesse können nicht unabhängig von dem Körper erfasst werden, der sie realisiert.

Die Abhängigkeit der Erklärung kognitiver Prozesse von ihrer physischen Realisierung ist zugleich ein Lösungsansatz für ein grundlegendes Problem funktionalistischer Ansätze im Sinne der Physical Symbol Systems Hypothesis oder der Language of Thought. Das Symbol-Grounding-Problem thematisiert die Frage nach der Verknüpfung der mentalen Symbole mit der Außenwelt. In Ansätzen, die das kognitive System als ein abstrakt-algorithmisches Verarbeitungssystem konzeptualisiert, das über diskreten Symbolen operiert, wird es zumeist entweder trivialisiert oder als peripherer nicht-kognitiver Gegenstand betrachtet und übergangen. Dabei stellt es ein zentrales Problem für diese Ansätze dar: Erklärt man bedeutungshafte Zustände in einem konzeptuellen System ohne Referenz auf die Außenwelt und ohne Interaktion mit dieser, so erhält man ein rein selbstreferenzielles System, in dem Bedeutung nur in Zirkelschlüssen verankert ist, denn die zur Definition eines Symbols herangezogenen Symbole sind selbst bedeutungslos und verweisen wiederum auf andere bedeutungslose Symbole. Diese Problematik wurde in verschiedenen Gedankenexperimenten verdeutlicht, etwa dem Chinese Room Argument (Searle, 1980), dem Chinese Dictionary Argument (Harnad, 1990) oder dem Irish Room Problem (McKevitt & Guo, 1996). Eine parallele Kritik an der kontextfreien, an der klassischen Künstlichen Intelligenz (KI) orientierten Erklärung von Kognition ohne Berücksichtigung ihrer Körpergebundenheit und Kontextualisierung äußert Dreyfus (1979).<sup>3</sup> Die Verknüpfung von Welt und abstraktem Symbol ist alles andere als trivial, denn es besteht keine eindeutige Beziehung zwischen den Strukturen, die hier aufeinander abgebildet werden.<sup>4</sup> Um dem selbstreferenziellen Kreislauf zu entkommen, muss das System aber zumindest Grundbedeutungen in seiner Beziehung zur Welt verankern – d.h. eine Verknüpfung zwischen Symbolen und den durch sie referierten Weltausschnitt etablieren. Dabei sind mit Symbol Grounding und Transduction zwei Aspekte zu unterscheiden: Zum einen wie die Symbole auf die Welt verweisen, zum anderen wie die Symbole ihrerseits bei einem entsprechenden Zustand der Außenwelt aktiviert werden. Wir müssen also beide Wirkungsrichtungen berücksichtigen, von der Welt zum Symbol und vom Symbol zur Welt. Damit zusammenhängend besteht noch die Entwicklungsfrage, wie ein Weltausschnitt oder eine Interaktion im kognitiven System in ein Symbol übertragen wird, d.h. wie die Symbole und wie ihre Verknüpfungen mit der Welt entstehen. In den Ansätzen der Embodied Cognition existiert das Problem des Symbol Grounding nicht. Man kann dabei allerdings nicht behaupten, dass das Problem in diesen Ansätzen gelöst wurde, vielmehr wird es hier umgangen. Sym-

<sup>3</sup>Die Kritik von Dreyfus (1979) erfolgt in der Tradition der phänomenologischen Philosophie Heideggers. Er bezieht sich dabei auf die Unmöglichkeit der Berechnung von Handlungsmöglichkeiten in einer gegebenen Situation auf Grundlage propositionaler Repräsentationen, da mit diesen nicht der holistische Bedeutungskontext der Dinge in der Welt erfasst wird (und auch nicht erfasst werden kann).

<sup>4</sup>Dies wird deutlich, wenn man das analoge Problem der Beziehung von Sprache auf die Welt betrachtet, in dem wir ebenfalls syntaktisch reglementierte, abstrakte Symbole auf Zustände der Welt abbilden. Erstens gibt es zumeist mehrere Weltzustände, die auf dasselbe Symbol verweisen. Zweitens können unterschiedliche Symbole und Symbolkombinationen auf denselben Weltzustand referieren. Drittens zeichnet sich die kategorielle Zuordnung von Welt durch unscharfe Grenzen und graduelle Zugehörigkeit aus (vgl. z.B. Rosch, 1977; Rosch & Lloyd, 1978; Mangasser-Wahl, 2000). Diskussionen zu Hindernissen einer einfachen Verknüpfung von Symbol und Welt finden sich unter anderem bei Lakoff (1987), Harnad (1990) und Glenberg (1997).

bole sind – zumindest in den Ansätzen der starken Embodied-Cognition-Hypothese – ein Korrelat unserer Erfahrungen mit der Welt, sei es mit unserer Wahrnehmung von Welt oder unserer Interaktion mit ihr. Dadurch müssen konkrete Weltausschnitte nicht mit abstrakten Symbolen verknüpft werden, die gänzlich anderen Zusammenhangsstrukturen folgen. Vielmehr sind Symbole direkte und konkrete Repräsentationen der Weltausschnitte und bewahren deren Zusammenhangsstrukturen. Das Grounding ist also keine nachträgliche Verknüpfung zwischen Symbol und Welt, sondern die Voraussetzung dafür, Symbole zu etablieren.

### 8.1.2. Konzeptualisierung und Kognition

Wählt man die Perspektive der Embodied Cognition als Ausgangspunkt für ein Erklärungsmodell menschlicher Kognition, ergibt sich eine Reihe grundlegender Fragen: Welcher Natur sind kognitive Repräsentationen? Wie sind diese strukturiert und wie erwerben wir sie? Wie können wir sie abrufen und wie nutzen wir sie für Prozesse des Denkens und Handelns? Den aktuellen Ansätzen liegt die Annahme zugrunde, dass die Prozesse und Strukturen unseres Denkens nicht unabhängig von ihrer Implementierung – d.h. dem menschlichen Körper mit seinen Eigenschaften und Modalitäten und seiner Einbettung in eine Umwelt – verstanden werden können. Dabei stehen unsere mentalen Repräsentationen in einem analogen Verhältnis zu der Umwelt, die sie repräsentieren. Denkprozesse sind dabei stark von den Prozessen unserer Wahrnehmung und unserer (motorischen) Handlung geprägt, d.h. von unserer Interaktion mit der Umwelt. Damit erhalten die Einflüsse der Modalitäten, über die das kognitive System mit der Umwelt interagiert, eine starke Gewichtung bei der Erklärung kognitiver Prozesse und Strukturen. Körper und Umwelt wirken als Constraints auf unsere kognitive Funktionen und ihre Entwicklung.

In verschiedenen theoretischen Ansätzen wurde versucht, auf dieser Basis ein funktionales konzeptuelles System zu modellieren, das Grundannahmen über unsere kognitiven Fähigkeiten erfüllen kann. In der aktuellen Forschung erfahren die folgenden drei Ansätze herausgehobene Popularität und wurden in verschiedenen Theorien zu Kognition und zum Sprachverstehen umgesetzt:

- Sensomotorische Erfahrungskorrelate für Handlungsmöglichkeiten als Bedeutungsträger; Affordances, Meshing und Suppression (Glenberg, 1997), Indexical Hypothesis (Glenberg & Robertson, 2000), Action-Sentence Compatibility Effect (Glenberg & Kaschak, 2002), Embodied Construction Grammar (Bergen & Chang, 2003; Bergen, Chang & Narayan, 2004).
- Simulation von Korrelaten der Perzeption als Bedeutungsträger; Perceptual Symbol Systems (Barsalou, 1999), Enactment und Simulation (Barsalou, Simmons, Barbey & Wilson, 2003), Simulationssemantik (Bergen, 2007), unmittelbare Aktivierung von Objekt-Eigenschaften (Zwaan, Stanfield & Yaxley, 2002; Stanfield & Zwaan, 2001), Immersed Experiencer Framework (Zwaan, 2004).
- Metaphorische Verankerung komplexer Bedeutung in einfachen räumlich-physischen Bedeutungen; Conceptual Metaphor (Lakoff, 1987; Lakoff & Johnson, 1999) und sprachliche Relativität (Casasanto, 2007; Casasanto & Boroditsky, 2008).

In den folgenden Abschnitten fassen wir zentrale Ideen dieser Ansätze zusammen. Im nächsten Unterkapitel gehen wir auf die theoretischen Ansätze zum Sprachverstehen ein, die daraus abgeleitet oder weiterentwickelt wurden.

### 8.1.2.1. Komplexe Konzepte durch Conceptual Metaphors

Eine frühe Einführung der Embodiment-Idee in die Psychologie erfolgt in der Kognitiven Linguistik (u.a. Lakoff, 1987; Langacker, 1987). In ihr werden abstrakt-syntaktischen Ansätzen des Sprachverstehens körpergebundene, erfahrungsbasierte, funktionale Alternativen gegenübergestellt. Besonders die Ausrichtung der Kognitiven Grammatik nach Lakoff legt einen Schwerpunkt auf Aspekte der Embodied Cognition. Ein zentrales Konzept sind dabei die „Conceptual Metaphors“, mit denen metaphorisches Denken als zentraler Mechanismus kognitiver Verarbeitung beschrieben wird (Lakoff & Johnson, 1980; Lakoff, 1987; Lakoff & Johnson, 1999). Aspekte unserer Körperlichkeit und unserer direkten wahrnehmungsbezogenen Erfahrungen stellen die Grundkonzepte, anhand derer ein Transfer zu komplexeren Konzepten geleistet wird. Dabei sind abstrakte und komplexe Konzepte sowie Emotionen in körperlichem und situationsbezogenem Wissen verankert (vgl. auch Barsalou, 1999) und werden durch metaphorische Prozesse gebildet. Conceptual Metaphors werden dabei nicht als linguistischer, sondern als kognitiver Gegenstand begriffen, d.h. in der beobachtbaren sprachlichen Verwendung von Metaphern spiegelt sich nur die kognitive Verwendung des metaphorischen Denkens wider.<sup>5</sup> Alle höheren kognitiven Prozesse beruhen dabei auf den konkreten körpergebundenen Grundkonzepten und werden durch deren Eigenschaften geprägt und beschränkt.<sup>6</sup>

Sprachliche Metaphern werden als Spiegel kognitiver Prozesse und Strukturen aufgefasst. Um diese Annahme zu stützen, muss geklärt werden, ob diese Relation tatsächlich besteht oder ob sprachliche Metaphern eigenen Gesetzmäßigkeiten folgen, die unabhängig von kognitiven Prozessen bestehen. In der Tat rückt die Annahme einer direkten Beziehung den Ansatz hin zu Fragen sprachlicher Relativität (vgl. auch Rohrer, 2001): In welcher Beziehung stehen Sprache und Denken zueinander und inwiefern ist Denken durch Sprache determiniert (zur aktuellen Debatte vgl. u.a. Levinson, 1996; Boroditsky, 2001; Boroditsky, Ramscar & Frank, 2001; Levinson, 2003; Boroditsky, 2003; Papafragou, 2005)? Bezogen auf die Annahme der Conceptual Metaphors stellt sich umgekehrt die Frage, inwiefern wir aus sprachlichen Relationen auf allgemeine kognitive Strukturen schließen können. Neben eher linguistischer und auf Intuition beruhender Erforschung der Conceptual Metaphors untersuchen Casasanto und Kollegen die Beziehung sowohl komplexer sprachlicher als auch komplexer kognitiver Konzepte (u.a. Emotionen, soziale Kategorien, zeitliche Relationen) zu primitiven kognitiven Konzepten körperlicher Wahrnehmung und Motorik und einfacher räumlicher Relationen (Casasanto, 2006, 2007; Casasanto & Boroditsky, 2008). Vor allem für die Konzeption von Zeit finden sie mehrfach Hinweise auf eine starke

<sup>5</sup>Zwar beschreiben Conceptual Metaphors kognitive und nicht sprachliche Strukturen, dennoch kann die sprachliche Verwendung von Metaphern und deren Verstehen wiederum genau über deren Begründung in Conceptual Metaphors untersucht und erfasst werden. Damit werden sie auch zu einem Werkzeug linguistischer Untersuchung. Eine solche Perspektive vertritt zum Beispiel Gibbs (2006).

<sup>6</sup>Kritik an den frühen Arbeiten Lakoffs betrifft zum einen die aus der Sicht der Gütekriterien empirischer Forschung (für eine Zusammenfassung vgl. z.B. Schmitt, 2001). Auf der anderen Seite wird mitunter die Vagheit und Inkonsistenz der Theorie bemängelt.

Verankerung in Konzepten wahrnehmbarer räumlicher Eigenschaften. Dabei können sich kulturbedingt Effekte unterschiedlicher Gerichtetheit der zugrundeliegenden räumlichen Konzepte auf die komplexen bzw. abstrakten Konzepte ergeben (vgl. Santiago, Lupianez, Pérez & Funes, 2007).

### 8.1.2.2. Konzeptualisierung über sensomotorische Handlungsmuster

In seinem Artikel „What memory is for“ kritisiert Glenberg (1997) an der bisherigen Gedächtnisforschung, dass sie sich in der Regel auf die Untersuchung der Gedächtniskapazität und der Merkleistungen für irrelevante, abstrakte Stimuli beschränkt, anstatt nach der Funktion unseres Gedächtnissystems zu fragen und unsere Gedächtnisleistung vor dem Hintergrund der Aufgaben zu untersuchen, für die sich dieses entwickelt hat. Die abstrakte Konzeptualisierung des Gedächtnissystems kann jedoch das Symbol-Grounding-Problem nicht umgehen, und blendet weiterhin den Einfluss der natürlichen Eigenschaften der Stimuli auf die Gedächtnisleistungen aus. Glenberg wählt für seine Gegenkonzeption aus der Perspektive der Embodied Cognition einen sehr weiten Gedächtnisbegriff, der das gesamte System kognitiver Konzeptualisierung umfasst. Dabei charakterisiert er Kognition als Konzeptualisierungsfähigkeit, die sich evolutionär im Dienste der Handlungsfähigkeit in einer realen dreidimensionalen Welt entwickelt hat, in der wir um unser Überleben zu sichern mit Gefahren umgehen und in Echtzeit reagieren müssen. Das perzeptuelle System hat sich entwickelt, um die Interaktion mit dieser realen Welt zu erleichtern.

Die Konzeptualisierung der Welt ist im kognitiven System in Mustern möglicher körperlicher Interaktionen mit dieser Welt realisiert. Diese sensomotorischen Handlungsmuster sind die eigentlichen Träger konzeptueller Bedeutung. Unser Gedächtnis ermöglicht es uns, unsere Interaktionserfahrungen und situationsbedingte Differenzierungen der ausführbaren Handlungsmöglichkeiten zu speichern. Unser Körper legt die Affordanzen zugrunde – die Handlungsmöglichkeiten in einer konkreten Situation mit den Eigenschaften und Fähigkeiten des uns gegebenen Körpers. Der Prozess der Konzeptualisierung erfolgt über das Meshing, der Integration von Handlungsmöglichkeiten und Handlungserfahrungen. In diesem funktionalen Constraint-Satisfaction-Prozess werden die Affordanzen für offen zugängliche Umwelteigenschaften kombiniert mit ähnlichen Aktionsmustern aus unseren früheren Erfahrungen, die im Gedächtnis gespeichert sind. Das resultierende integrierte Handlungsmuster ist die Konzeptualisierung, die ein Individuum in einer spezifischen Situation entwickelt. Bedeutung entspricht also einem personalisierten Muster möglicher Interaktionen zwischen Körper und Welt. Genau dadurch wird das erworbene Wissen handlungsrelevant und unterstützt aus evolutionärer Perspektive unsere Lebensfähigkeit. Es bleibt zu hinterfragen, ob eine solche Einschätzung von Bedeutung nicht massiv beschränkt ist und eine große Klasse kategorisier- und verstehbarer Objekte und Ereignisse ausschließen, die keine eigene Handlung der verstehenden Person einschließen.

Glenberg (1997) geht davon aus, dass dasselbe Gedächtnissystem, das für die Kategorisierung und für die Berechnung von Handlungsmöglichkeiten zuständig ist, auch dem Erinnerungsvermögen unseres semantischen und episodischen Gedächtnisses zugrunde liegt. Die Ausführung von Erinnerungsprozessen, erinnerungsbasiertem Lernen und abstrakten Denkprozessen, die unabhängig vom tatsächlichen Umweltzustand stattfinden – etwa die Planung und die Antizipation möglicher Handlungen – erfordern aber, dass die

mentalen Zustände und Simulationen von Handlungen und Erinnerungen nicht starr mit ihrer Wahrnehmung, ihrer sensomotorischen Umsetzung und tatsächlichen Ausführung gekoppelt sind. Glenberg nimmt hierfür die Existenz eines Suppression-Mechanismus an, der es dem System ermöglicht, diese Entkopplung der Denkprozesse zu erreichen und die tatsächliche Handlungsausführung zu unterdrücken. Suppression ermöglicht die Funktion des Gedächtnisses in zwei verschiedenen Modalitäten: (a) Eine automatische Verarbeitung ohne Intention, die unbewusst, online und umweltbezogen erfolgt, und (b) eine bemühte Verarbeitung, die intentional, offline und bewusst erfolgt. Letztere ist inhärent rekonstruktiv und basiert auf Suppression bzw. Entkopplung. Suppression wird also zur Grundlage höherer kognitiver Prozesse. Welcher Natur dieser Mechanismus ist, wie er angesteuert wird und wie er auf andere Prozesse einwirkt, bleibt leider unspezifiziert.

### 8.1.2.3. Konzeptualisierung und Simulation von Perzeption

Barsalou (1999) schlägt mit den „Perceptual Symbol Systems“ einen ähnlichen Ansatz zur kognitiven Konzeptualisierung vor, der aber stärker an der Perzeption orientiert ist. Auch hier sind analoge Korrelate der Erfahrung mit der Umwelt der Kern der Bedeutung, und höhere kognitive Prozesse sind als Simulationen konzeptioniert. Barsalou's Ansatz entsteht aus der Kritik an abstrakt-amodalen Symboltheorien der klassischen KI-orientierten Kognitionswissenschaft und der mit ihr einhergehenden Verbannung von bildhaften (bzw. allgemeiner formuliert zur repräsentierten Umwelt analogen perzeptionserhaltenden) Wissensrepräsentationen. In einer Charakterisierung abstrakt-amodaler Symboltheorien erläutert er Probleme und Schwächen. Dabei thematisiert er neben der allgemeinen Abwesenheit empirischer Evidenzen für die Existenz diskreter abstrakter Symbole vor allem die Frage des Symbol Grounding und der Transduktion. In diesem Zusammenhang kritisiert er die unbegründete fundamentale Distinktion zwischen Perzeption und Motorik (als Peripherie der Kognition) gegenüber einer vom sensomotorischen System unabhängig operierenden Kognition. Da auf Grundlage abstrakt-amodaler Symbolsysteme aber viele fundamentale Prinzipien und Leistungsmerkmale menschlicher Kognition erfolgreich beschrieben und modelliert werden können, ist das erklärte Ziel, ein auf der Perzeption begründetes, voll-funktionales konzeptuelles System zu entwickeln.

In Barsalou's Gegenkonzeption erfolgt kognitive Konzeptualisierung anhand perzeptueller Symbole, die in der Wahrnehmung verankert sind. Diese nutzen als Korrelat der körpergebundenen Wahrnehmung dieselben Ressourcen und Repräsentationen, auf denen unser Wahrnehmungsapparat operiert (vgl. Barsalou, 1999; Barsalou et al., 2003; Barsalou, 2008). Oder auf der neurofunktionalen Erklärungsebene betrachtet: Perzeptuelle Symbole sind Korrelate neuronaler Zustände, die während einer Wahrnehmung oder während einer Handlung aktiviert waren, und durch unsere perzeptuellen Erfahrungen geprägt werden. Kognitive Prozesse, die auf diesen perzeptuellen Symbolen operieren, finden folglich in denselben Arealen statt, in denen sie während unserer Erfahrungen entstanden sind. Aus dieser Grundeigenschaft folgen mehrere Eigenschaften des Systems: (a) Symbole, auf denen höhere kognitive Prozesse operieren, erben Eigenschaften unserer Perzeption, (b) Symbole sind nicht amodal, sondern multimodaler Natur, (c) Symbole sind nicht abstrakt, sondern analog insofern, als sie die strukturellen und inhaltlichen Relationen des Gegen-

stands erhalten, den sie abbilden, und (d) Kognition und Perzeption lassen sich nicht mehr trennen, sie sind vielmehr als ein System zu begreifen.

Barsalou leitet eine Reihe von Eigenschaften her, die ein solches Symbolsystem sowie dessen Grundeinheit (die perzeptuellen Symbole) charakterisieren und aus denen die Leistungsfähigkeit für ein voll-funktionales konzeptuelles System resultieren. Die Symbole entstehen in einem schematischen Symbolformationsprozess, der auf selektiver Aufmerksamkeit und isolierter Speicherung beruht. Deshalb sind die Symbole eine Untergruppe kohärenter Aspekte einer Wahrnehmung und als solche ebenfalls schematisch und selektiv. Wäre dies nicht der Fall, wären sie nur individuelle holistische kontextabhängige Speicherungen von Wahrnehmungen, über die das System keine Kategorisierungen leisten könnte und für die es keine Kompositionalität besitzen würde. Stattdessen sind perzeptuelle Symbole allgemein-generalisierend, komponentiell, (bedingt) kontextunabhängig und kategorial. Entgegen abstrakten diskreten Symbolen sind perzeptuelle Symbole aufgrund ihrer Prägung und Veränderlichkeit durch Erfahrung dynamisch. Nicht zuletzt sind sie multimodal, d.h. sie sind sowohl im visuellen, auditiven, haptisch, olfaktorisch-gustativen Wahrnehmungssystem sowie in Propriozeption und Introspektion verankert. Dabei entstehen die Symbole in den jeweiligen Modalitäten nach denselben Prinzipien und aus denselben Mechanismen. Über assoziative Areale werden zudem übergeordnete multimodale Kategorien gebildet.

Wahrnehmungsunabhängige, höhere kognitive Prozesse erfolgen als Simulationen von Wahrnehmungssituationen auf Grundlage der perzeptuellen Symbole, die für die Simulation in sogenannten Simulatoren organisiert werden. Auf Grundlage dieser spezifizierten Simulatoren können Simulationen für räumlich und zeitlich abwesende Gegenstände und Situationen und gleichermaßen für Planung und Antizipation sowie hypothetische und kreativ-kontrafaktische Ereignisse und Gegenstände generiert werden. Dafür werden Untermengen von Symbolen in sogenannten Frames aktiviert und integriert. Die Simulationen sind also skizzenhaft, beliebig konkret oder abstrakt und manipulierbar, und können folglich gleichermaßen individuelle, variierte oder prototypische Instanzen generieren. Aufgrund der Schematizität der symbolischen Repräsentationen erfolgen die Aktivierungen unbewusst – wobei sie aber bewusste Gegenstücke haben können. Auf neurofunktionaler Erklärungsebene können Simulationen als Re-Enactment der neuronalen Aktivität von Wahrnehmungsereignissen erfasst werden.

Die ad hoc-Kategorisierung von Umweltereignissen entspricht einem erfolgreichen Abgleich zwischen einer generierten Simulation und der Wahrnehmung des entsprechenden Ausschnitts der Umwelt. Durch die Eigenschaften der Kategorisierungsmechanismen und den Rückgriff auf beliebig reiche multimodale Informationen können darüber hinaus kategoriale Inferenzen sowie Brückeninferenzen, Antizipation und Interpretation geleistet werden. Die Eigenschaften dieses einheitlichen Systems von Repräsentationsmechanismen ermöglichen seine multiple Funktionalität. Die Eigenschaften der Simulationen über den Frames ermöglichen weitere grundlegende Eigenschaften: Die Bildung von Prädikaten, Attribut-Wert-Bindungen und Propositionen, sowie den Einbezug von Constraints und rekursiven Strukturen. Mit der Darstellung der Entstehung von solchen symbolischen Verarbeitungseigenschaften demonstriert Barsalou (1999) die prinzipielle Möglichkeit, in einem perzeptuell verankerten System volle konzeptuelle Funktionalität realisieren zu können. Allerdings bleibt die Theorie mitunter sehr vage. Vor allem die Herleitung

komplexer und abstrakter Begriffe (angelehnt an Conceptual Metaphors, vgl. Lakoff, 1987) sowie propositionaler Strukturen bleibt bisweilen wenig überzeugend, und ist als ad hoc hergeleitete Erklärung kaum für eine Umsetzung und Überprüfung in empirischen Studien handhabbar. Unklar ist auch, weshalb Barsalou auf der Verwendung symbolischer Strukturen beharrt, deren Merkmal es ja gerade ist, diskret und abstrakt zu sein.

### 8.1.2.4. Koexistenz verschiedener Wissensstrukturen

In der jüngeren Forschung wurden verschiedentlich Relativierungen der starken Embodied-Cognition-Hypothese hervorgebracht, die die Rolle körperlich verankerter Symbole verringern oder eine Koexistenz analog-modaler und amodal-abstrakter Symbole annehmen. Louwrese (2007) schlägt in seiner Hypothese der Symbol-Interdependenz vor, dass Symbole ein interdependentes abstraktes System der Bedeutung bilden, das partiell in der Erfahrung des Verstehenden mit der Welt ge-groundet ist. Er stellt damit die Annahme in Frage, dass wirklich jedes Wort ge-groundet sein muss, um verstanden werden zu können. Als Alternative schlägt er im Sinne der Symbolic-Theft-Hypothese (Cangelosi, Greco & Harnad, 2002) vor, dass abstrakt-amodale Symbole auf analog-modale Symbole gesattelt werden können: Sobald grundlegende situational verankerte Konzepte entwickelt sind, können aus diesen neue, kontextunabhängige abstrakt-amodale Symbole kombiniert werden (zu einer ähnlichen Argumentation kultureller Vermittlung vgl. Bryson, 2005). In der Language-And-Situated-Simulation-Theorie (LASS-Theorie, Barsalou, Santos, Simon & Wilson, 2008) wird zunächst ähnlich argumentiert. Sie nimmt die Koexistenz verschiedener Wissenssysteme an, unter anderem das Situated-Simulation-System sowie das auf Kookkurrenz linguistischer Erfahrungen basierende linguistische System. Die Repräsentationen des linguistischen Systems sind allerdings explizit nicht amodale Symbole, sondern modale Korrelate früherer Erfahrungen mit Sprache. Im Unterschied zu Louwrese (2007) und Cangelosi et al. (2002) nehmen Barsalou et al. (2008) nicht an, dass die linguistischen Formen bedeutungshafte Repräsentationen bilden können, oder dass über Kombinationsprozesse aus einfachen verkörperlichten Konzepten abstrakte, von der körperlichen Erfahrung unabhängige Konzepte gebildet werden können. Tiefgehendes konzeptuelles Wissen ist ausschließlich im Situated Simulation-System verfügbar.

## 8.2. Sprachverarbeitung im Embodiment-Ansatz

Gerade im Forschungsschnittpunkt Sprachverstehen mit seinen vielfältigen Anknüpfungspunkten – etwa Wahrnehmung, Gedächtnis, Sprachverarbeitung, Konzeptualisierung und Kategorisierung, Sequenzialisierung, Produktivität, Inferenzen – gibt es eine wachsende Zahl von Ansätzen, die auf Grundlage der Annahmen der Embodied Cognition theoretische Modelle des Sprachverstehens konzipieren. In diesem Zusammenhang wird mitunter von den Forschungsfeldern Embodied Language Processing und von Simulationssemantik gesprochen. Die hierbei getroffenen Annahmen über die zugrundeliegenden Prinzipien wurden anhand empirischer Studien mit Methoden aus Psychologie, Linguistik, Psycholinguistik und Neurowissenschaften untermauert und spezifiziert.

Die theoretische Grundlage des Ansatzes ist die Prägung des kognitiven Systems durch die Eigenschaften des Körpers und seiner Modalitäten sowie seiner Situiertheit in einer Umwelt. Die Konzeptualisierung des Sprachverstehens erfolgt parallel zum Verstehen von Umweltsituationen in diesem System. Sprache funktioniert dabei als Surrogat tatsächlich wahrgenommener Erfahrungen: Sprachverstehen entspricht der Aktivierung oder Konstruktion von Konzeptualisierungen der sprachlich beschriebenen Situationen. Da der Erfassung der Bedeutung sprachlicher Äußerungen sensomotorische Repräsentationen zugrunde liegen, werden Interaktionen zwischen diesen Verarbeitungsmodalitäten erwartet. Es wurden verschiedene Ansätze zum Verhältnis zwischen Embodied Cognition und Sprachverarbeitung ausgearbeitet, die den Zugriff von Sprache auf die als perzeptuelle und motorische Korrelate konzeptualisierte Bedeutung beschreiben und Konsequenzen daraus ableiten. Wir besprechen kurz einige zentrale Konzeptionen und Rahmentheorien:

- Indizierung von konzeptualisierten Handlungsmustern durch sprachliche Zeichen (Barsalou, 1999; Glenberg & Robertson, 2000; Glenberg & Kaschak, 2002)
- Simulationssemantik als Simulation von zur sprachlichen Äußerung korrespondierenden Wahrnehmungs- und Handlungsaspekten (Barsalou, 1999; Glenberg & Robertson, 2000; Rohrer, 2005; Bergen, 2007), Suppression zur Gewährleistung der wahrnehmungs- und handlungsunabhängigen Verarbeitung von Sprache (Glenberg, 1997)
- Conceptual Metaphors als Brücke zu komplexen und abstrakten Begriffen und zur Erfassung zugrundeliegender atomarer Bedeutungskomponenten (Lakoff & Johnson, 1980; Casasanto & Boroditsky, 2008)
- Rahmentheorien und Formalisierungen für linguistische und sprachpsychologische Ansätze mit der Embodied Construction Grammar (Bergen & Chang, 2003) und dem Immersed Experiencer Framework (Zwaan, 2004)

### 8.2.1. Linguistische Indizierung von Konzepten

Im Rahmen der verschiedenen Ansätze eines in körperlicher Erfahrung verankerten konzeptuellen Systems wurden recht ähnliche Vorschläge gemacht, wie die Verknüpfung von Sprache mit der eigentlichen Bedeutung – den nicht-sprachlichen sensomotorisch gegroundeten Konzepten – erfolgt und wie Sprachverstehen etabliert werden könnte. Explizit formuliert wurde ein Vorschlag mit der Indexical Hypothesis (Glenberg & Robertson, 2000, S. 379): „[...] the meaning of a sentence is constructed by (a) indexing words and phrases to real objects or perceptual, analog symbols; (b) deriving affordances from the objects and symbols; and (c) meshing the affordances under the guidance of syntax.“ Mit der zweiten und dritten Stufe der Bedeutungskonstruktion werden wir uns im nächsten Abschnitt befassen. Zunächst werfen wir aber einen kurzen Blick auf deren Grundlage: Um mit sprachlichem Input, d.h. linguistischen Einheiten, sensomotorisch-erfahrungsbasierte Bedeutungen zu aktivieren, müssen erstere mit letzterem in irgendeiner Weise verbunden sein. Glenberg und Robertson (2000) erläutern diesen Aspekt leider nicht sehr ausführlich. Es wird nur davon gesprochen, dass Wörter und Phrasen auf perzeptuelle Symbole (im Sinne von Barsalou, 1999) indiziert oder abgebildet werden – ein Wort referiert auf ein

Objekt der Außenwelt bzw. auf eine prototypisch-modale mentale Repräsentation dieses Objekts (vgl. auch Kaschak & Glenberg, 2000; Glenberg & Kaschak, 2002).

Etwas genauer erläutert wird die Indizierung und deren Entwicklung im Spracherwerb durch Barsalou (1999). Linguistische Symbole entwickeln sich gemeinsam mit den mit ihnen assoziierten perzeptuellen Symbolen. Sie sind selbst ebenfalls modal, d.h. sie entwickeln sich nach denselben Mechanismen wie perzeptuelle Symbole über selektive Aufmerksamkeit als schematische Korrelate körperlicher Erfahrungen – in diesem Fall Erfahrungen aus der Wahrnehmung und Handlung mit gesprochener und geschriebener Sprache. Und die Sprachkorrelate werden ebenfalls organisiert über Simulatoren in Simulationen dieser linguistischen Erfahrungen, um Wörter und Phrasen zu erkennen und zu produzieren. Während ihrer Entstehung werden die linguistischen Simulatoren mit den Simulatoren der Erfahrungskorrelate assoziiert, auf die sie referieren. Die assoziierten Simulatoren können gleichermaßen ganze Objekte, Teilobjekte, Objekteigenschaften, Relationen oder Events sein. Die sensomotorisch-konzeptionellen Simulatoren werden hinsichtlich ihrer verschiedenen Aspekte über die Zeit mit vielen verschiedenen linguistischen Simulatoren assoziiert – dabei wird ein semantisches Feld generiert, das das zugrundeliegende konzeptuelle Feld spiegelt (vgl. Barsalou, 1999, S. 592). Nach ihrer assoziativen Verknüpfung können die sensomotorischen, d.h. die eigentlich bedeutungstragenden Simulatoren durch die linguistischen Simulationen angesteuert bzw. ausgelöst werden. Die linguistischen Simulatoren ermöglichen also die Kontrolle der Simulation von Wahrnehmungsereignissen: Das Verstehen von Sprache in Text und Diskurs entspricht einer durch Lexik und Syntax geleiteten Konstruktion einer korrespondierenden Simulation des sprachlich beschriebenen Ereignisses.

Bei dieser Konzeption der Indizierung von Wörtern auf perzeptuelle Symbole fällt allerdings auf, dass dieselbe Trivialisierung betrieben wird, die im Rahmen des Symbol-Grounding-Problems hinsichtlich der Verbindung abstrakt-arbiträrer Symbole mit Weltzuständen und -ereignissen kritisiert wird. Denn eigentlich sollte die Indizierung genau denselben Problemen gegenüberstehen: Erfordert sie in ihrer bisherigen Charakterisierung nicht ebenfalls Grounding- und Transduktionsprozesse? Treten bei der Indizierung von linguistischen auf perzeptuell-motorische Konzepte nicht ebenfalls Zuweisungsprobleme durch fehlende Eineindeutigkeit (Synonymie, Polysemie, Homonymie), unscharfe Grenzen und graduelle Zugehörigkeit (Prototypikalität) auf? So betrachtet fehlt in diesem theoretischen Ansatz eine Spezifizierung, wie diese Indizierung genau funktionieren soll. Damit bleibt Barsalou eine Erklärung schuldig, weshalb die Indizierung ein nebensächliches Problem sein sollte.

### 8.2.2. Linguistisch kontrollierte Simulation

Nach der linguistischen Indizierung sensomotorischer Konzepte beim Sprachverstehen folgen laut der Indexical Hypothesis (Glenberg & Robertson, 2000) zwei weitere Schritte. Beim zweiten Schritt werden die indizierten Konzepte genutzt, um Affordanzen – die durch die Eigenschaften des Körpers vorgegebene und durch Interaktionserfahrung geprägte Handlungsmöglichkeiten in einem situativen Kontext – abzuleiten. Im dritten Schritt schließlich werden die abgeleiteten Affordanzen aus den Konzepten durch die linguistischen Informationen in einem Meshing-Prozess kombiniert bzw. integriert. Die

Bedeutung eines Satzes entspricht der sprachlich angeleiteten mentalen Handlungskonstruktion aus den aktivierten Handlungsmöglichkeiten – Sprache wird durch die kognitive Simulation der vom Satz implizierten Handlungen bedeutungsvoll gemacht. Das Resultat des Meshings entspricht damit konzeptuell den Simulationen bei Barsalou (1999). Das erfolgreiche Verstehen des Satzes setzt voraus, dass der Verstehende aus seiner Erfahrung die entsprechenden Affordanzen ableiten und in der Folge das Meshing erfolgreich erfüllen kann.

Glenberg und Robertson (2000) schreiben dem Indexical Hypothesis-Ansatz Eigenschaften zu, die ihn von symbolischen und dabei vor allem intensionalen Bedeutungstheorien unterscheiden (vgl. auch Glenberg & Kaschak, 2002). So konstatieren sie, dass durch die Verankerung der bedeutungstragenden Repräsentationen in situieren körperlichen Interaktionen das Symbol-Grounding-Problem umgangen wird. Die mentale Repräsentation von sprachlicher Bedeutung ist nicht sprachlicher, sondern situational-handlungsbezogener Natur. In dieser Hinsicht hat die Indexical Hypothesis Bezüge zur Theorie Mentaler Modelle (vgl. etwa Johnson-Laird, 1980, 2005). Als Voraussetzung für das Verstehen von Sprache im Sinne einer Simulation identifiziert Glenberg (1997) die Suppression. Um sprachliche Äußerungen verstehen zu können, die sich nicht auf aktuelle Aspekte der Wahrnehmung beziehen (wenn z.B. jemand von seinem Rom-Urlaub im vergangenen Jahr erzählt), müssen die für das Verstehen dieser Äußerung stattfindenden kognitiven Prozesse von der tatsächlichen sensorischen Wahrnehmung und motorischer Handlung im Hier und Jetzt entkoppelt werden. Wie bereits erwähnt wird allerdings nicht erklärt, wie diese Suppression funktionieren soll.

Während der Ansatz von Glenberg und Robertson (2000) auf dem Begriff der perzeptuellen Symbole aufbaut und Verstehen ebenfalls als Simulation konzipiert, unterscheidet sich der Ansatz von Barsalou (1999) dadurch, dass er Bedeutung weniger in der auf Affordanzen beruhenden Simulation von Handlungen realisiert sieht, sondern auf Simulationen perzeptueller Erfahrungen im Sinne eines modalen Erlebens der beschriebenen Situation. Durch die Indizierung der Simulatoren durch linguistische Pointer können nun gezielt Simulationen ausgelöst werden. Sprache bzw. die durch sie übertragene grammatische Information wird in diesem Ansatz zu einer Konstruktionsanweisung. Sie dient als kontrollierendes Element, das die Konstruktion und Integration entsprechender Simulatoren leitet. Sprachverstehen entspricht der erfolgreichen Konstruktion einer Simulation, die zur sprachlich beschriebenen Situation korrespondiert. Umgekehrt werden bei der Sprachproduktion durch konzeptuelle Simulationen assoziierte linguistische Simulatoren aktiviert, die als Kandidaten für zu äußernde Wörter und syntaktische Strukturen dienen. Damit wird Sprache zum Medium für die Koordination von Simulationen über verschiedenen Individuen hinweg.

Mit dem Begriff der Simulationssemantik fasst Bergen (2007) die verschiedenen Ansätze zusammen, in denen Sprachverstehen als Simulation konzipiert wird (vgl. auch Bergen, 2005; Bergen & Chang, 2003). Simulationssemantik steht als Überbegriff für Forschungsansätze, die verschiedene Aspekte der Sprache in ihrem Einfluss auf die Spezifik der mentalen Simulation untersuchen. Mit ihr soll geklärt werden, was in welcher Weise simuliert wird und wie genau die Simulationen durch sprachliche Mittel angesteuert, kontrolliert und spezifiziert werden. Nach wie vor problematisch ist die theoretische Erfassung von Simulationen komplexer und abstrakter Gegebenheiten und Ereignisse sowie die Erklärung

von Negation.<sup>7</sup> Grundlage der Erfassung von Abstrakta ist der Ansatz der Conceptual Metaphors (Lakoff, 1987; Lakoff & Johnson, 1999, vgl. Kapitel 8.1.2.1), dessen Umsetzung im Ansatz der Simulationssemantik als komplexe Abfolge von und Interaktion zwischen Simulationen formuliert wird.

### 8.2.3. Perspektivische Situationssimulation

Es wurden verschiedene Rahmenkonzeptionen entwickelt, die den Prozess des Sprachverstehens aus Perspektive der Embodied Cognition weiter spezifizieren und aus sprachwissenschaftlicher Perspektive nutzbarer machen, indem sie sie in eine übergreifende allgemeine Theorie des Sprachverstehens integrieren. Ein sehr prominenter Ansatz ist das Immersed Experiencer Framework (IEF, Zwaan, 2004), das die Beziehungen zwischen verschiedenen linguistischen Ebenen und Simulationsaspekten weiter ausführt. Hinsichtlich der angenommenen ablaufenden Prozesse weist der Ansatz vielfach Parallelen zu anderen psychologischen Ansätzen des Sprachverstehens auf, so etwa zum Construction-Integration-Modell (Kintsch, 1998; Kintsch & van Dijk, 1978), zum Structure Building Framework (Gernsbacher, 1990), zum Constructivist Framework (Graesser et al., 1994) oder zum Event-Indexing-Modell (Zwaan & Radvansky, 1998). Allerdings zielt das Immersed Experiencer Framework darauf ab, Probleme amodal-abstrakter propositionaler Sprachverstehenstheorien aufzugreifen und zu überwinden.<sup>8</sup> Dabei soll es einen Rahmen bieten, in den empirische Befunde pro Embodied Cognition und Embodied Language Processing in die Theoriebildung integriert werden können. Darüber hinaus soll der Ansatz die Ableitung weiterer und vor allem spezifischerer Hypothesen ermöglichen.

Zwaan (2004) greift in seinem Ansatz die Idee auf, dass Sprachverstehen einer mentalen Simulation der sprachlich beschriebenen Situation entspricht, und führt diese weiter. Die Grundidee ist dabei, dass der Verstehende einer sprachlichen Äußerung die damit beschriebene Situation als ein in die Situation eingetauchter Erlebender erfährt – der Verstehende simuliert mental das sensomotorische Erleben der sprachlich beschriebenen Situation, das nachempfundene Erleben entspricht dem erfolgreichen Verstehen der Äußerung.<sup>9</sup> Grundla-

---

<sup>7</sup> Gerade der Aspekt der Negation ist schwer in einer auf erfahrungsbasierten Simulationen begründeten Sprachverstehenstheorie umzusetzen, weil der negierte Sachverhalt ja nicht einfach nicht simuliert werden kann, da damit der Gegenstand der Negation nicht spezifiziert wäre. In einem Vorschlag hierzu wird angenommen, dass Negation als gestufter Prozess erfolgt, bei dem zuerst der negierte Sachverhalt simuliert wird. Erst in einem zweiten Schritt erfolgt die Negation durch einen Aufmerksamkeitswechsel auf den tatsächlichen Sachverhalt (vgl. Kaup & Zwaan, 2003; Kaup, Lüttke & Zwaan, 2005, 2006; Kaup, 2006; Kaup, Yaxley, Lüttke & Zwaan, 2007).

<sup>8</sup> Zwaan (2004) diskutiert auch empirische Befunde, die für eine propositionale Repräsentation sprechen. Dem stellt er die Behauptung gegenüber, dass das Immersed Experiencer Framework dieselben Befunde vorhersagt, darüber hinaus aber präzisere und spezifischere Vorhersagen trifft, die auf Basis von Propositionen nicht erfasst werden könnten. In dieser Hinsicht charakterisiert er Propositionen als Kurzform wissenschaftlicher Beschreibungen der zugrundeliegenden Strukturen, die nur über eine eingeschränkte Vorhersagequalität verfügen, und deshalb keinen Ersatz für die detailliertere und präzisere Langform wissenschaftlicher Beschreibungen darstellen, die durch das Immersed Experiencer Framework gegeben ist.

<sup>9</sup> Eine wichtige Rolle spielt in diesem Ansatz die Perspektive, die bei der Simulation des Erlebens auf die Situation eingenommen wird. Dieser Fokus auf die Rolle der Perspektive wird im Ansatz von MacWhinney (2008) weiter vertieft, in dem Grundlagen der Grammatik aus der Wahl der Perspektive bei der sensomotorisch verankerten Sprachverarbeitung abgeleitet werden.

ge dafür ist die Aktivierung von Korrelaten perzeptueller und motorischer Erfahrungen mit der Umwelt (im Sinne von Barsalou, 1999; Glenberg & Robertson, 2000). Das Ziel des Verstehensprozesses ist die inkrementelle Konstruktion einer mentalen Repräsentation der Situation. Dabei werden im Immersed Experiencer Framework drei Hauptkomponenten dieses Prozesses unterschieden: Activation, Construal und Integration. Diese erfolgen während der Verarbeitung als zeitlich überlappende Interpretationsereignisse, die auf verschiedenen Sprach- und Situationsebenen operieren (vgl. Abb. 8.1).

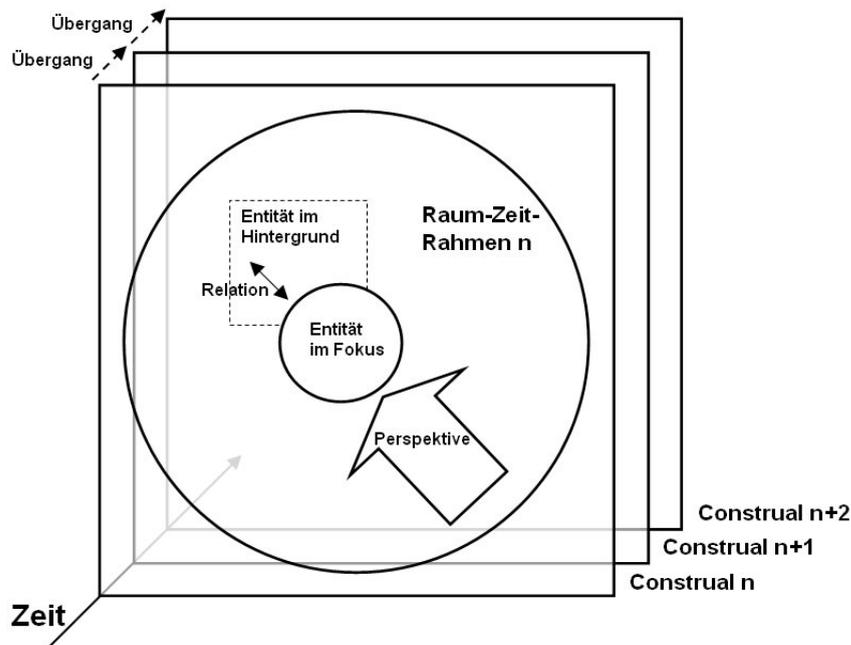


Abbildung 8.1.: Schematische Skizze der Hauptkomponenten des Immersed Experiencer Framework nach Zwaan (2004, S. 42), Abbildung in leicht modifizierter Version.

Activation beschreibt den Aktivierungsprozess von funktionalen Aktivierungsmustern, die über den ganzen Kortex verteilt sind. Er erfolgt auf Wortebene. Der Prozess resultiert in einer diffusen Aktivierung überlappender Repräsentationen verschiedener Instantiierungen des assoziierten Referenten, wobei der Grad der Diffusion erfahrungsabhängig ist und etwa durch Frequenz, Primat und Rezenz geprägt wird. Die aktuelle Aktivierung stellt den Kontext nachfolgender Aktivierungen funktionaler Aktivierungsmuster – der Grad der Spezifiziertheit bestimmt die Beschränkung möglicher Folgeaktivierungen.

Construal beschreibt den Constraint-Satisfaction-Prozess, über den funktionale Aktivierungsmuster in eine gemeinsame Interpretation integriert werden. Durch die gegenseitigen Beschränkungen werden die Aktivierungsmuster im größeren Kontext spezifiziert. Der Prozess erfolgt auf Phrasen- und Satzebene und steuert die mentale Simulation von einzelnen Situationen bzw. Ereignissen. Sprachverstehen dient hier als sprachbasierte Modulation der Aufmerksamkeit auf einer situativen Gegebenheit. Construals sind aufgrund der gerichteten Aufmerksamkeit und der Verarbeitungsökonomie schematisch. Ihre Bildung erfolgt unmittelbar und inkrementell, und spezifiziert sowohl neu hinzukommende als auch vorangegangene Bedeutungsaspekte. Als Situationsmodell umfassen sie vier zentrale Komponenten: Den zeitlichen Rahmen, die räumliche Verortung, die Perspektive

auf die Situation, und die Entitäten (Objekte und Personen im Fokus und im Hintergrund des Ereignisses) mit ihren Eigenschaften und Relationen zueinander. Dabei werden die Simulationen beim Verstehen von Sprache durch die Erfahrungen geprägt, die wir in der tatsächlichen Wahrnehmung von Objekten, Raum, Zeit und Perspektive gesammelt haben.

Integration beschreibt Übergänge zwischen Construals, über die Veränderungen und Entwicklungen der beschriebenen Situation sowie Übergänge zu neuen Situationen erfasst werden. Dieser Prozess erfolgt auf Satz- und Text- bzw. Diskursebene. Auch die mit diesem Prozess verbundenen Eigenschaften der Simulationen sind erfahrungsbasiert, wobei Übergänge, die nicht unseren Gewohnheiten und Erfahrungen entsprechen, schwerer zu verarbeiten sind. Die Komplexität der Integration wird dabei von vier Faktoren bestimmt: (1) der Erwartung von Kontinuität, d.h. der Erhaltung des räumlichen, zeitlichen und perspektivischen Rahmens, (2) der Erwartung überlappender Informationen, d.h. Situationsveränderungen müssen sukzessive erfolgen, nur einzelne Komponenten dürfen jeweils verändert werden, (3) der Vorhersagbarkeit im Sinne der Antizipierbarkeit kommender Ereignisse, etwa durch kausale Relationen oder Planstrukturen, sowie (4) der sprachlichen Hinweise auf einen Situationsübergang.

### 8.2.4. Embodied Construction Grammar

Eine zweite Rahmenkonzeption, die kurz erwähnt werden soll, ist die Embodied Construction Grammar (Bergen & Chang, 2003; Bergen et al., 2004). Den verschiedenen Ansätzen der Konstruktionsgrammatik<sup>10</sup> liegen die Annahmen zugrunde, dass nicht syntaktische Regeln, sondern Form-Bedeutungs-Paare die zentrale Komponente sprachlichen Wissens darstellen, sowie dass Syntax und Lexikon keine getrennten Ressourcen, sondern beide verschieden komplexe Realisierungen linguistischer Form sind. Konstruktionen sind also keine Strukturanweisungen für kleinere sprachliche Einheiten, sondern selbst beliebig große Minimaleinheiten, die über Konstruktionsfamilien organisiert sind, und jeweils eine eigene funktionale Bedeutung besitzen (für eine Übersicht zu zentralen Konzepten der Konstruktionsgrammatik vgl. z.B. Evans & Green, 2006; Goldberg, 2006; Tomasello, 2006). Mit der Embodied Construction Grammar wird nun versucht, eine Verbindung zwischen den Ansätzen der Konstruktionsgrammatik und der Simulationssemantik zu formulieren und damit eine grammatische Formalisierung des Embodied Language Processing zu etablieren, über die formale linguistische Analysen und die Erstellung komputationaler Modelle ermöglicht werden.

In der Embodied Construction Grammar ist linguistisches Wissen in grundlegenden Paar-Einheiten von Form und Bedeutung organisiert, wobei die Bedeutungskomponenten als detaillierte sensomotorische Repräsentationen angenommen werden. Diese Repräsentationen sind in verkörperlicht-erfahrungsbasierten Schemata organisiert. Konstruktionen sind in diesem Ansatz Strukturen, die mentale Simulationen im sensomotorischen System parametrisieren. Der Verstehensprozess besteht aus den zwei Subprozessen Analyse und Simulation: In der Analysephase werden Konstruktionen im linguistischen Input identifiziert und im System instantiiert, in der Simulationsphase erfolgt das eigentliche Verstehen

---

<sup>10</sup>Die Konstruktionsgrammatik ist kein einheitlicher Ansatz grammatischer Formalisierung, sondern eher ein Forschungsfeld, in dem sich verschiedene Realisierungen etabliert haben, so etwa die Construction Grammar, die Radical Construction Grammar oder die Cognitive Grammar.

der Bedeutung als Simulation entsprechend der durch die Konstruktion spezifizierten Parameter.

Insgesamt wirkt der Ansatz von Bergen und Chang (2003) unausgereift: Gerade die interessanten Aspekte werden in der Formalisierung übergangen und bleiben theoretisch un-spezifiziert. So bleiben Aspekte an der Schnittstelle zwischen Konstruktionsgrammatik und Embodied Cognition vage und die Annahme von Simulationen zur Bedeutungskonstruktion wirkt künstlich aufgesetzt.<sup>11</sup> So ist kein wirklicher Unterschied zu propositionalen Theorien erkennbar, denn die Formalisierung sieht explizite Role-Filler-Kodierungen mit aufgesetzter sensomotorischer Ansteuerung von Schemata vor – allesamt klassisch-symbolische Konzepte – womit darüber hinaus das Symbol-Grounding-Problem wieder eingeführt wird. Es stellt sich die Frage, ob sich Aspekte der Embodied Cognition, etwa Simulation und Constraint-Satisfaction-Prozesse, auf diese Art und Weise formalisieren lassen. Denn wenn durch die Formalisierung Eigenschaften der Sprachverarbeitung wiederum über abstrakt-arbiträr diskrete (und nur bedingt modale) Strukturen erfasst werden, werden eben gerade die zentralen Voraussetzungen aufgegeben, die mit der Embodied Cognition etabliert werden sollen.

### 8.3. Empirische Befunde zum Embodiment-Ansatz

Zur Erforschung der Zusammenhänge zwischen perzeptueller und motorischer Erfahrung auf der einen und Sprachverarbeitung sowie Gedächtnis und Inferenzprozesse auf der anderen Seite haben sich verschiedene Zugangsweisen etabliert. Diese unterscheiden sich hinsichtlich der untersuchten Dimension des Verstehens (Motorik, Perzeption, Abstrakta, Situationsmodelle) und der zur Untersuchung verwendeten Methode. Vorwiegend werden anhand psycholinguistischer Experimentalparadigmen behaviorale Daten erhoben. Dabei spielen vor allem Priming-Studien eine große Rolle, mit denen Effekte der Kompatibilität und der Interferenz von Stimuli in verschiedenen Modalitäten untersucht werden, um den Bezug zwischen perzeptuell-motorischen Repräsentationen und Sprachverstehensprozessen zu erfassen. Auch Gedächtnisexperimente werden verwendet, anhand derer der Einfluss situationaler analog-modaler Repräsentationen auf die Präsenz von Gedächtnisinhalten bemessen werden kann. In neurophysiologischen Studien werden anhand bildgebender Verfahren (v.a. fMRI, PET und ERP) Gemeinsamkeiten und Unterschiede neurofunktionaler Korrelate von sprachlichen und perzeptuellen Stimuli erhoben. Im Zentrum steht der Versuch, die Bedeutung allgemeiner Prozesse sensomotorischer Verarbeitung für das Verstehen sprachlicher Äußerungen zu untersuchen. Erfolgt Sprachverstehen tatsächlich über die Simulation allgemeiner kognitiver Prozesse, so sollten sich Wechselwirkungen mit diesen zeigen lassen. Wir stellen in den folgenden Absätzen zentrale Ansätze und Ergebnisse vor.

---

<sup>11</sup>Ähnliche Schwächen entsprechender Versuche zu Formalisierung zeigen sich auch bei der Exemplifizierung komplexer Simulationen bei Barsalou (1999).

### 8.3.1. Action-Sentence Compatibility Effect

Die im Forschungsfeld momentan vermutlich populärste Studie, in der die Interaktion zwischen Sprachverstehen und allgemeiner sensomotorischer Verarbeitung untersucht wird, stammt von Glenberg und Kaschak (2002). In dieser wird eine aus den Annahmen der Indexical Hypothesis (Glenberg & Kaschak, 2002) abgeleitete empirische Hypothese überprüft: Mit der Indexical Hypothesis wird vorgeschlagen, dass Sprache dadurch bedeutungsvoll wird, dass die in Sätzen implizierte Handlung mental simuliert wird. Folglich müsste die sprachlich induzierte simulierte Handlung mit tatsächlich ausgeführten motorischen Handlungen interferieren, da sie auf dem gleichen Substrat bzw. über dieselben Ressourcen operieren.

Zur Überprüfung wurden Grammatikalitätsbeurteilungs- bzw. Priming-Experimente durchgeführt. In diesen wurden Experimentalsätze in zwei Bedingungen manipuliert, die wir anhand Beispiel 29 erläutern (Beispielsätze nach Glenberg & Kaschak, 2002): Die Sätze implizieren eine Bewegung, die entweder zum Körper eines Referenten des Satzes hin oder vom Körper dieses Referenten weg gerichtet ist. So impliziert der Imperativsatz 29a *Öffne die Schublade* eine zum Referenten her gerichtete Bewegung, während der Imperativsatz 29b *Schließe die Schublade* eine vom Referenten weg gerichtete Bewegungsausführung impliziert. Zudem wurde der Satztyp variiert: Neben Imperativsätzen wurden konkrete (Bsp. 29c und 29d) und abstrakte (Bsp. 29e und 29f) Transfersätze verwendet, die jeweils wie im Beispiel in Doppel-Objekt- und in Dativ-Struktur realisiert wurden.

- (29) a. Open the drawer. (*Zum Körper hin gerichtet*)  
 b. Close the drawer. (*Vom Körper weg gerichtet*)  
 c. Andy handed you the pizza. (*Zum Körper hin gerichtet*)  
 d. You handed Andy the pizza. (*Vom Körper weg gerichtet*)  
 e. Lizz told you the story. (*Zum Körper hin gerichtet*)  
 f. You told Lizz the story. (*Vom Körper weg gerichtet*)

Die Sätze wurden visuell Wort für Wort auf einem Bildschirm im RSVP-Paradigma<sup>12</sup> dargeboten. Aufgabe der VTn war es, die Sätze nach ihrer Grammatikalität (grammatisch oder ungrammatisch) zu beurteilen, wofür das Experimentalmaterial entsprechend um Nonsense-Sätze erweitert wurde. Zur Beurteilung mussten sie Knöpfe auf einer Buttonbox betätigen: Zur Präsentation der Sätze musste ein mittlerer Knopf gedrückt gehalten werden, zur anschließenden Beurteilung musste ein Knopf gedrückt werden, der entweder näher zum VTn oder weiter entfernt von diesem auf der Buttonbox liegt, wobei ein Knopf für eine positive und der gegenüberliegende für eine negative Beurteilung zugewiesen wurde. Die Zuweisung wechselte nach der Hälfte des Experiments. Aufgezeichnet wurde die Latenz zwischen dem Satzende und dem Lösen der zentralen Taste.

Sollte Sprachverstehen die mentale Simulation einer Handlung bewirken, so sollte diese Simulation mit einer entsprechenden tatsächlichen Handlungsausführung interagieren.

<sup>12</sup>Rapid Serial Visual Presentation: In diesem Paradigma hat der VTn im Unterschied zum Self-Paced Reading keinen Einfluss auf die Darbietungszeit einzelner Wörter. Die Darbietung erfolgt nach einer durch den Versuchsleiter vorab festgelegten Taktung. Mit dieser Methode können Antworten bzw. Entscheidungen und deren Latenzen erfasst werden.

Die Ausführung einer tatsächlichen Handlung ist also einfacher, wenn sie zur sprachlich beschriebenen Handlung kompatibel ist. Umgekehrt bedeutet dies aber auch, dass die Inkompatibilität von sprachlich beschriebener und tatsächlich auszuführender Handlung zu Interferenzen führen muss. Dies sollte sich im Experiment darin widerspiegeln, dass das Drücken der Knöpfe in kompatibler Richtung schneller erfolgt als in inkompatibler Richtung.<sup>13</sup>

Die Ergebnisse bestätigen eine Interaktion zwischen sprachlich implizierter und tatsächlich ausgeführter Handlungsrichtung in den Reaktionszeiten.<sup>14</sup> Diese Interaktion wird als Action-Sentence Compatibility Effect (ACE, Glenberg & Kaschak, 2002) bezeichnet. Der Befund des Effekts für konkrete Transfersätze weist darauf hin, dass keine eigentliche Aufforderung zur Handlungsausführung enthalten sein muss, um eine entsprechende motorische Simulation zu erzeugen. Der Befund des Effekts für abstrakte Transfersätze bestätigt darüber hinaus die Annahme der Conceptual-Metaphor-Theorie: Die Repräsentation abstrakter Gegenstände und Handlungen scheint in motorischen Repräsentationen verankert zu sein oder beinhaltet diese zumindest.<sup>15</sup>

Die ACE-Studie wurde vielfach repliziert und mit verschiedenen Manipulationen spezifiziert, wobei unter anderem das sprachliche Material, die Modalität des dargebotenen Stimulusmaterials, die Abstraktionsebene der Sätze, die für die Reaktion erforderliche Bewegung, die Darbietungszeit und der Darbietungszeitpunkt für die Reaktion variiert wurden (vgl. u.a. Bergen & Wheeler, 2009; Bergen, 2005; Borreggine & Kaschak, 2006). So können etwa Borreggine und Kaschak (2006) zeigen, dass der Kompatibilitätseffekt vom Timing der Handlungsplanung abhängig ist: Können die VTn die motorische Handlungsausführung für die Beurteilungsaufgabe bereits während der Verarbeitung des Satzes planen, findet sich der ACE. Wird die Antwortrichtung allerdings erst nach Offset des jeweiligen Stimulussatzes vorgegeben, verschwindet der Effekt.

In verschiedenen Studien zum ACE zeigen sich nicht nur Kompatibilitätseffekte, sondern auch Interferenz-Effekte, d.h. bei Kompatibilität zwischen sprachlichem Stimulus und motorischer Reaktion zeigt sich eine Erschwerung der Handlungsausführung (z.B. Bergen, 2005). Diese genau entgegengesetzte Reaktion wird als Überlagerung der übereinstimmend aktivierten Repräsentationen erklärt. Leider wurde bislang keine zufriedenstellende Erklärung dafür erbracht, weshalb sich die Effekte in einen Fall als Voraktivierung (und folglich als Erleichterung) auswirken und im anderen Fall als Überlagerung (und folg-

---

<sup>13</sup>Hierzu ein Beispiel: Wenn der VTn den Satz *Open the drawer.* hört (impliziert eine Bewegung zum Körper hin) und zur Beurteilung dieses Satzes als grammatikalisch einen Knopf drücken muss, der eine Bewegung hin zum Körper fordert (kompatible Bedingung), so sollte diese Bewegung erleichtert sein und schneller bzw. früher erfolgen. Wenn der Knopf zur Beurteilung als grammatikalisch allerdings eine Bewegung weg vom Körper fordert (inkompatible Bedingung), so sollte die Ausführung erschwert sein.

<sup>14</sup>In einem Kontrollexperiment wurde die Beurteilungsaufgabe so gestaltet, dass keine Bewegung mit gerichteter Handlungsausführung erforderlich war (die Beurteilung erfolgte mit der rechten und der linken Hand auf einer horizontal liegenden Buttonbox). Tatsächlich wurden die Effekte in dieser Variante eliminiert, was auf eine Abhängigkeit von der Ausführung einer kompatiblen oder inkompatiblen gerichteten Bewegung hinweist.

<sup>15</sup>So impliziert das Erzählen einer Geschichte (Bsp. 29e und 29f) keine eigentliche motorische Bewegung, sondern nur eine Handlungsrichtung im metaphorischen Sinne. Diese implizite metaphorische Handlungsrichtung genügt aber, um die tatsächliche Ausführung einer kompatiblen oder inkompatiblen Handlung zu beeinflussen. Dies weist darauf hin, dass die motorische Repräsentation zumindest ein Bestandteil der Repräsentation des abstrakten Sachverhalts (des Erzählens) sein muss.

lich als Erschwerung). Ebenfalls ist unklar, durch welche Faktoren bedingt wird, ob ein Kompatibilitätseffekt oder ein Interferenz-Effekt erfolgt. Eine Diskussion dieses Sachverhalts erfolgt durch Bergen (2007), der die Effektrichtung auf die zeitliche Relation der Stimuli, der Aufgabe und der motorischen Planung zurückführt – synchrone Stimulus-Response-Verknüpfung führt eher zu Überlagerung (im Sinne einer Konkurrenz um die Verarbeitungsressourcen), sukzessive Stimulus-Response-Verknüpfung eher zu Voraktivierung. Doch auch diese Argumentation bleibt vage und kann nur bedingt präzise Vorhersagen ermöglichen.

#### 8.3.2. Effekte der Perzeption: Form, Farbe, Orientierung

Während die Studie von Glenberg und Kaschak (2002) Aspekte motorischer Simulation behandelt, beschäftigt sich eine weitere, mittlerweile klassische Studie mit dem Gegenstück der perzeptuellen Simulation: Zwaan et al. (2002) untersuchen die Hypothese, dass das Verstehen von Sprache einer Simulation perzeptueller Erfahrung entspricht. Sie nähern sich dieser Annahme über die Frage, ob der Verstehende einer sprachlichen Äußerung grundsätzlich und unmittelbar Detailinformationen der darin referierten Objekte referieren, etwa ihre Oberflächenform, ihre Farbe oder ihre Orientierung im Raum (vgl. auch Stanfield & Zwaan, 2001; Zwaan & Yaxley, 2003; Dahan & Tanenhaus, 2005).

Zwaan et al. (2002) verwenden hierfür Priming-Experimente mit Picture-Matching-Task.<sup>16</sup> Als Stimulus verwenden sie Sätze, die für einen dort erwähnten Referenten eine bestimmte äußere visuell distinktive Form nahelegen. So legt etwa Satz 30a nahe, dass der Adler ausgebreitete Flügel hat, da er sich in der Luft befindet. Satz 30b hingegen legt einen Adler mit angelegten Flügeln nahe, da er sich im Nest befindet. Aufgabe der VTn war es, den jeweiligen Stimulussatz zu lesen, und für ein anschließend auf dem Bildschirm präsentiertes Bild zu bestimmen, ob das darauf abgebildete Objekt im zuvor gelesenen Satz erwähnt wurde. Das präsentierte Bild zeigte entweder eine zur implizierten visuellen Form kongruente Abbildung des Referenten oder eine inkongruente Abbildung, zum Beispiel für Satz 30a einen Adler mit ausgebreiteten oder im inkongruenten Fall mit eingezogenen Flügeln. Die Beurteilung erfolgte per Tastendruck (Experiment 1), aufgezeichnet wurde die Korrektheit und die Latenz der Beurteilung. In Experiment 2 erfolgte eine verbale Benennung des abgebildeten Referenten anstatt eines Tastendrucks. Mit dieser Variante war damit kein direkter Vergleich der Korrespondenz der Referenten in Satz und Bild erforderlich.

- (30) a. The eagle is in the sky.  
b. The eagle is in the nest.

Die Ergebnisse zeigen einen hochsignifikanten Effekt der Kongruenz in den Latenzen: Bedingungen mit Kongruenz zwischen Bild und Satz wurden deutlich schneller bestätigt

---

<sup>16</sup>Im Picture-Matching-Task werden den VTn Bilder und Sätze oder Wörter präsentiert (zeitlich versetzt oder gleichzeitig, über die gleiche Modalität oder cross-modal über verschiedene Modalitäten). Die VTn haben die Aufgabe zu beurteilen, wie gut das Bild zum gegebenen Satz bzw. Wort passt oder das Wort bzw. der Satz zum gegebenen Bild. Dabei kann sowohl die Antwort als auch im Sinne einer Priming-Studie die Latenz für die Beurteilung aufgezeichnet werden.

als inkongruente Bedingungen. Auch für die Korrektheit zeigten sich Tendenzen eines Kongruenzeffekts. Experiment 2 bestätigte diese Befunde. Die Ergebnisse weisen auf eine sehr frühe Aktivierung von Objekteigenschaften sprachlich benannter Objekte, ohne dass diese Eigenschaften explizit benannt werden. Sie zeigen auch, dass der Satzkontext unmittelbar die Repräsentation der Objekte beeinflusst. Dies legt nahe, dass die Aktivierung der Objekteigenschaften eher impliziter Natur sind und durch den im Satz beschriebenen situationalen Kontext spezifiziert werden.

Eine methodisch übereinstimmende Studie von Stanfield und Zwaan (2001) belegt, dass die sprachlich induzierte Repräsentation benannter Objekte nicht nur in ihrer äußeren Form, sondern auch hinsichtlich ihrer Orientierung im Raum bestimmt ist. Nach Sätzen, die für ein Objekt eine spezifische Orientierung nahelegen, werden schematische Abbildungen dieser Objekte schneller erkannt, wenn diese in einer kongruenten Orientierung dargeboten werden. So wird nach Satz 31a ein waagrecht abgebildeter Nagel schneller erkannt als ein senkrecht abgebildeter, da ein Nagel unserer Erfahrung nach in horizontaler Ausrichtung in die Wand geschlagen wird. Entsprechend verhält sich das Erkennen der Abbildungen für Satz 31b genau umgekehrt. Weitergehende Ergebnisse zeigen, dass die angenommene Konstruktion des mentalen visuellen Modells nicht nur für die Orientierung von Objekten im Raum sensitiv ist, sondern auch für ihre Bewegungsachse (Richardson, Spivey, Barsalou & McRae, 2003) und ihre Bewegungsrichtung auf den Bewegungsachsen (Bergen, Lindsay, Matlock & Narayanan, 2007).

- (31) a. He pounded the nail into the wall.  
b. He pounded the nail into the floor.

Die Unmittelbarkeit der Aktivierung von Informationen, die die visuelle Form von Objekten betreffen, wurde in einer alternativen Herangehensweise unter anderem von Dahan und Tanenhaus (2005) untersucht. Anstatt eines Priming-Experiments verwenden sie eine Eyetracking-Studie im Visual-World-Paradigma. In dieser bilden sie auf einem Bildschirm neben einem Zielobjekt (z.B. einer SCHLANGE) verschiedene Distraktorobjekte ab, von denen eines eine figurale Ähnlichkeit zum Zielobjekt aufweist (z.B. ein eingerolltes SEIL) und zwei nicht (z.B. ein SOFA und ein REGENSCHIRM). Die Aufgabe der VTn war es, jenes Objekt anzuklicken, das per auditiver Darbietung benannt wurde (übereinstimmend zum Zielobjekt, in unserem Beispiel das Wort *Schlange*). Die Ergebnisse zeigen bereits während der Nennung (d.h. vor Off-Set des Wortes) eine starke Ablenkung der Fixationen durch das figural ähnliche Distraktor-Objekt. Über eine Rating-Studie wurde sichergestellt, dass die Ähnlichkeit nicht durch eine rein oberflächliche visuelle Ähnlichkeit der Objekte zustandekommt, und folglich tatsächlich auf die figurale Ähnlichkeit zurückzuführen ist.<sup>17</sup> Zudem wurden die jeweils zusammen dargestellten Objekte so gewählt, dass keine phonetischen Ähnlichkeiten zwischen diesen auftreten. Die Ergebnisse legen nahe, dass der Wort-Objekt-Abgleich auf der Ebene lexikalisch aktivierter perzeptueller Merkmale stattfindet. Zudem weisen sie darauf hin, dass routinemäßig und unmittelbar die Form eines Objekts mitaktiviert wird.

---

<sup>17</sup>So wurde die SCHLANGE im Beispiel zum Beispiel ausgestreckt und das SEIL eingerollt dargestellt. Die rein oberflächliche Ähnlichkeit war also gering (visuelle Ähnlichkeit). Seile erinnern aber hinsichtlich ihrer Form generell an Schlangen (figurale Ähnlichkeit).

Weitergehende Untersuchungen führte unter anderem Bergen (2005) durch (vgl. auch Bergen et al., 2007), in denen es um die Spezifikation der Rolle der linguistischen Informationen bei der mentalen Simulation perzeptueller Features für bewegte Objekte geht: Welche linguistischen Elemente lösen die mentale Simulation aus oder spezifizieren sie? Genügen einzelne lexikalische Elemente für die Simulation oder wird ein Satz benötigt? Die Ergebnisse geben Hinweise darauf, dass die Konstruktion des mentalen visuellen Modells sowohl durch Objekt-Nomen als auch durch Hauptverben und deren stereotypische Orientierung und Bewegungsrichtung angesteuert werden. Zusätzliche Kontrollexperimente zeigen aber auch, dass die Simulation der Bewegungsrichtung von Objekten nicht alleine durch die Aktivierung der lexikalischen Elemente erzeugt, sondern in einer Simulation der Situation durch den gesamten Satz spezifiziert wird – Wörter formen Spezifikationen für Einzelaspekte in der Situation, Sätze steuern die Konstruktion des mentalen Situationsmodells.<sup>18</sup>

Verschiedentlich wurde argumentiert, dass die Annahme der Aktivierung solcher Zusatzinformationen, die für die Gesamtaussage eines Satzes irrelevant sind, nicht der Logik entspricht, die in symbolischen Theorien mit einer propositionalen Repräsentation der Satzbedeutung zugrundegelegt wird (vgl. z.B. Zwaan et al., 2002; Zwaan, 2004). Allerdings ist anzumerken, dass der Versuchsaufbau in den meisten Studien die Berücksichtigung visueller Informationen einfordert. Demzufolge besteht die Möglichkeit, dass die Aktivierung perzeptueller Informationen dadurch bedingt oder verstärkt wird, dass diese zur Erfüllung der gestellten Aufgabe (vorwiegend Picture-Matching) benötigt werden, ohne dass deshalb eine mentale visuelle Simulation angenommen werden muss.<sup>19</sup> Eine weitere Kritik betrifft die Frage, ob die Aktivierung von Korrelaten sensomotorischer Erfahrungen tatsächlich kausale und bedeutungstragende Elemente der kognitiven Verarbeitung sind oder nur ein Epiphänomen der eigentlichen Verarbeitung, die amodal und algorithmisch über abstrakten symbolischen Repräsentationen erfolgt (vgl. z.B. Barsalou, 2008). Die sensomotorischen Repräsentationen könnten also erst in einem Folgeschritt aufgrund ihrer starken Assoziation mitaktiviert werden und dabei keinerlei Rolle für die konzeptuelle Verarbeitung und das Verstehen eines Satzes spielen. Zwei häufig angeführte, aber die Diskussion keinesfalls abschließende Gegenargumente sind, (a) dass die Aktivierung sensomotorischer Merkmale unmittelbar erfolgt, und dieser minimale Zeitraum bis zur Aktivierung eine kausale Rolle dieser Repräsentationen nahelegt, und dass (b) bislang umgekehrt kaum empirische Evidenzen für abstrakt-amodale Repräsentationen erbracht wurden.

#### 8.3.3. Neuropsychologische Befunde

Neben behavioralen Studien wurden auch neurophysiologische Studien mit bildgebenden Verfahren eingesetzt. Diese unterstützen die Annahme der Embodied-Cognition-These, dass Bedeutungen Korrelate der neuronalen Repräsentation von tatsächlich erlebten Wahrnehmungen und ausgeführten Handlungen sind, indem sie nachweisen, dass sich Perception, Motorik und höhere kognitive Prozesse dieselben neurofunktionalen Ressourcen teilen

---

<sup>18</sup>In einer weiteren Studie erweitern Bergen und Wheeler (2009) diese Hypothese noch um den grammatischen Aspekt des Verbs (hier: progressive und perfect): Inhaltswörter bestimmen den Gegenstand der Simulation, während der Aspekt des Satzes bestimmt, wie die Simulation ausgeführt wird.

<sup>19</sup>Ähnliche Kritik wurde und wird an Studien zum Visual Imagery geübt (vgl. Pylyshyn, 2003).

bzw. die verwendeten Ressourcen korrespondieren. In verschiedenen Studien wurde dazu die neuroanatomische Verortung neuronaler Aktivierungen für tatsächlich ausgeführte Handlungen oder die Beobachtung tatsächlich ausgeführter Handlungen mit denen korrespondierender sprachlicher Äußerungen verglichen, die diese Handlungen benennen oder miteinbeziehen.<sup>20</sup> Eingesetzt werden vor allem funktionelle Magnetresonanztomographie (fMRT), Positronen-Emissions-Tomographie (PET) und ereigniskorrelierte Potentiale (EKP) im Elektroenzephalogramm.

Pulvermüller (1999) stellt die Aktivierung modalitätsspezifischer kortikaler Areale bei der Verarbeitung korrespondierender Wortklassen fest: Während die Aktivierung für die phonetische Form von Wörtern sowie hochabstrakte Wörter (Funktionswörter) stark lateralisiert und vor allem im perisylvischen Kortex lokalisiert sind, aktivieren Inhaltswörter Areale, die über beide Hemisphären verteilt sind. Wörter, die auf visuelle Stimuli verweisen, schließen die Aktivierung visueller Kortizes mit ein. Wörter, die auf Handlungen verweisen, aktivieren Areale in motorischen Kortizes. In einer späteren Studie, die sich konkret auf Hypothesen der Embodied Cognition bezieht, untersuchen Pulvermüller, Härle und Hummel (2001) anhand EKP und fMRT Aktivierungskorrelate für auditiv präsentierte Aktionsverben, die Handlungen mit unterschiedlichen Körperteilen beschreiben (*to talk* für Gesicht/Mund, *to write* für Arm/Hand, *to walk* für Bein/Fuß), und vergleichen sie mit Aktivierungskorrelaten der tatsächlichen Ausführung von Handlungen mit den Körperteilen, die zu diesen Wörtern korrespondierenden. Tatsächlich finden sie ab etwa 250 ms nach Wort-On-Set signifikante topographische Unterschiede der Hirnaktivität, die durch die unterschiedlichen Verben ausgelöst werden. Die Differenzen beziehen sich auf Areale des motorischen und des angrenzenden frontalen Kortex, mit stärkster Aktivierung von Regionen, in denen die entsprechenden Körperbereiche (Gesicht, Arm, Bein) repräsentiert werden, die primär für die Ausführung der jeweils beschriebenen Handlungen verwendet werden.

Zu ähnlichen Ergebnissen kommen Tettamanti et al. (2005) sowie Aziz-Zadeh, Wilson, Rizzolatti und Iacoboni (2006). Aziz-Zadeh und Kollegen vergleichen in einer fMRT-Studie Aktivierungskorrelate für Sätze, in denen Handlungen mit unterschiedlichen Effektoren (wiederum Gesicht, Arm, Bein) beschrieben werden, mit Aktivierungskorrelaten der visuellen Wahrnehmung korrespondierender Handlungen. Sie fanden ebenfalls eine Kongruenz zwischen diesen effektorspezifischen Aktivierungen der visuell und linguistisch präsentierten Stimuli. Eine interessante behaviorale Studie, die sich direkt auf die Befunde von Pulvermüller et al. (2001) bezieht, stellen Scorolli und Borghi (2007) vor: Sie untersuchen vergleichbar zur Studie von Glenberg und Kaschak (2002), ob sprachliche Stimuli das motorische System modulieren können. In ihrem Experiment mussten VTn effektorbezogene Nomen-Verb-Paare danach beurteilen, ob sie sinnvoll sind oder nicht. Die Beurteilung wurde mit den entsprechenden Effektoren durchgeführt (mündlich durch aussprechen des Urteils, mit dem Fuß durch das Drücken eines Fußpedals). Kongruente Effektoren führten zu geringeren Latenzen bei der Beurteilung. Dies weist auf sehr detaillierte Simulationen mit effektorspezifischem Anteil hin. Zudem unterstützen diese Ergebnisse die Befunde der

---

<sup>20</sup>Die meisten Studien wurden bislang speziell zu motorischen Korrelaten durchgeführt. Entsprechende neurofunktionale Untersuchungen für perzeptuelle Korrelate (etwa im Sinne von Zwaan et al., 2002) wurden kürzlich von Hirschfeld und Zwitterlood (2010) und Hirschfeld, Zwitterlood und Dobel (2010) vorgestellt.

neurofunktionalen Studien – sie zeigen, dass die gefundenen Aktivierungsunterschiede tatsächlich mit einem behavioral relevanten Verarbeitungsunterschied einhergehen.

So klar die Ergebnisse aus den neurofunktionalen Studien erscheinen mögen, so bleibt dennoch die Unklarheit bestehen, ob die Aktivierung von Korrelaten sensomotorischer Erfahrungen tatsächlich konzeptuell-bedeutungstragende Elemente höherer kognitiver Verarbeitungsprozesse darstellen oder nur Epiphänomene dieser Prozesse sind. Abgesehen davon beinhalten die Studien auch Befunde, die schwer mit den Ansätzen der starken Embodied-Cognition-These in Einklang zu bringen sind. So scheinen Funktionswörter keine erkennbare Verankerung in sensomotorischen Repräsentationen aufzuweisen (vgl. Pulvermüller, 1999). Dies lässt sich allgemein auf Abstrakta ausweiten: Bislang gibt es keine Befunde für eine Aktivierung sensomotorischer Areale bei der Verarbeitung abstrakter Wörter (vgl. z.B. Aziz-Zadeh et al., 2006). Dies deckt sich mit Befunden aus behavioralen Studien, die auf ein Ausbleiben entsprechender Effekte bei metaphorischer Sprache hinweisen (vgl. Bergen et al., 2007).

#### 8.3.4. Perspektivische Simulationen in der Satzverarbeitung

In unserer Forschungsgruppe im Psycholinguistischen Forschungslabor der Universität Freiburg wurde eine Reihe von Experimenten durchgeführt, die sich mit der Spezifizierung der Perspektive und der Gliederung des simulierten Raums bei der mentalen Simulation auseinandersetzen.<sup>21</sup> Grundlage der Studien ist die Annahme, dass beim Verarbeiten eines Satzes ein perspektivische Simulation eines korrespondierenden, räumlich und zeitlich verteilten Situationsmodells erfolgt (Glenberg & Kaschak, 2002; Zwaan, 2004; MacWhinney, 2008; Markman & Brendl, 2005). Hierzu wurden die Fragen gestellt, (a) welche linguistischen Aspekte die Wahl der Perspektive des Verstehenden innerhalb der simulierten Situation bestimmen, und (b) auf welche Weise sich diese Perspektive und die Gliederung des mentalen Raums auf die Verarbeitung motorisch verankerter Wortklassen auswirkt bzw. ob Anaphern zu dieser Klasse von Wörtern gehören.

Verschiedenen Theorien liegt die Annahme zugrunde, dass beim Verstehen von Sprache eine sensomotorische Simulation des Erlebens der sprachlich beschriebenen Situation generiert wird. In dieser Simulation nimmt der Verstehende eine spezifische Perspektive auf diese Situation ein. Welche Perspektive er aber einnimmt, durch welche linguistischen Hinweise diese moduliert wird und ob möglicherweise Präferenzen für die Perspektivenwahl bestehen, diese Details bleiben zum Teil unspezifiziert oder werden überhaupt nicht besprochen. So wird bei Zwaan (2004) standardmäßig die Perspektive eines außenstehenden, an der Handlung unbeteiligten Beobachters angenommen. Die Befunde von Glenberg und Kaschak (2002) und anderen, die eine Simulation gerichteter motorischer Handlungen nahelegen, deuten allerdings darauf hin, dass der Verstehende sich gegebenenfalls

---

<sup>21</sup>Die im folgenden vorgestellten Experimente wurden nicht vom Autor durchgeführt oder initiiert. Vielmehr war er mit weiteren Kollegen an der Planung, Vorbereitung und Durchführung sowie als Co-Autor bei den Veröffentlichungen beteiligt. Die Studie zu Embodied Cognition und Perspektivenwahl (Schwarzkopf, Weldle, Müller & Konieczny, 2011, accepted; Schwarzkopf et al., 2008) war Teil der Magisterarbeit von Sarah Schwarzkopf (Schwarzkopf, 2008), die Studie zu Embodied Cognition und Anaphernaufflösung (Konieczny & Weldle, 2009; Konieczny, Weldle, Haser et al., 2010; Konieczny, Haser et al., 2010) wurde maßgeblich von Lars Konieczny initiiert.

in die Rolle des Ausführenden hineinversetzt (was in Zwaans Modell allerdings nicht ausgeschlossen wird). Um die Rolle des Ausführenden einzunehmen, muss man sich in diesen hineinversetzen. Spezifischer wird diese Projektion bei der Perspektiveneinnahme und -übernahme von MacWhinney (2008) beschrieben. Sprachliche Grammatik reflektiert dabei die kognitiven Operationen, die der Perspektiven-Verfolgung zugrundeliegen – sie ist die Anleitung, wie wir im mentalen Modell der beschriebenen Situation die aktuelle Perspektive und Perspektivenwechsel nachvollziehen. Je nach gegebener grammatischer Information wird eine unterschiedliche Perspektive, etwa eine ego- oder allozentrische Perspektive, auf die Situation gewählt.

**Simulation und Perspektivenwahl** Verschiedene Aspekte des sprachlichen Stimulus kommen infrage, die die Perspektivenwahl bestimmen oder zumindest beeinflussen könnten, unter anderem die grammatische Person, die syntaktische Funktion, die thematische Rolle, die Nennungsreihenfolge, der Fokus oder die Salienz. Um dies zu untersuchen, verwenden Schwarzkopf et al. (2008) die Methodik der Studien von (Glenberg & Kaschak, 2002), und überprüfen eine Serie von Manipulationen von Sätzen, in denen einige dieser möglichen Determinanten der Perspektive variiert werden (vgl. auch Schwarzkopf, 2008; Schwarzkopf et al., 2011, accepted).<sup>22</sup> Der ACE wurde damit als Indikator für die Übernahme einer bestimmten Perspektive – als Sender oder Empfänger einer gerichteten Handlung – eingesetzt: Findet sich ein Kompatibilitätseffekt in eine Richtung, lässt sich darüber die eingenommene Perspektive ableiten. Die Effekte weisen auf eine Wahl der Perspektive hin, die durch die grammatische Person in Interaktion mit der syntaktischen Funktion bzw. der thematischen Rolle gewählt wird (diese beiden Aspekte können in diesem Experiment nicht unterschieden werden).<sup>23</sup> Dabei tritt nur dann ein ACE auf, wenn ein Referent in der 1. Person Singular erscheint. Je nachdem, in welcher Rolle bzw. Funktion dieser Referent erscheint, zeigt sich für ein Verb mit spezifischer Handlungsrichtung eine Erleichterung oder eine Erschwerung. Damit bestätigen die Ergebnisse auch die Befunde von Bergen (2005); Bergen et al. (2007); Bergen und Wheeler (2009): Der ACE ist nicht alleine von der impliziten Handlungsrichtung des Verbs (z.B. *schieben* immer vom Körper weg) bestimmt, sondern wird durch die Handlungsrichtung der Situation moduliert, die durch den Satz beschrieben wird.

Konvergierende Evidenzen für diese Befunde wurden in Experimenten von (Brunyé et al., 2009) mit Picture-Matching-Task gezeigt. In diesen mussten die VTn die Übereinstimmung von Sätzen mit Satzsubjekt in der 1., 2. oder 3. Person Singular mit Bildern beurteilen, die die beschriebene Handlung aus einer egozentrischen oder einer allozentrischen Perspektive darstellen. Für die 1. und 2. Person Singular wurde die egozentrische Perspektive schneller als passend beurteilt, für die 3. Person Singular hingegen die außenstehende Beobachterperspektive. Allerdings zeigte ein zweites Experiment, dass diese Projektion durch den Diskurskontext manipuliert und umgekehrt werden kann, wenn dieser Kontext die Projektion in den Handelnden blockiert oder zur dispräferierten Version macht – etwa

---

<sup>22</sup>In einem der Experimente wurden Subjekt und Objekt der Stimulussätze so variiert, dass (a) die 1. Person Singular als Agens und Subjekt (gegenüber einem Referenten in der 3. Person Singular), (b) die 1. Person Singular als Patiens und Objekt (gegenüber einem Referenten in der 3. Person Singular) oder (c) beide Referenten in der 3. Person Singular erscheinen.

<sup>23</sup>Den Ergebnissen muss allerdings mit einiger Vorsicht begegnet werden, da in den ANOVAs zumeist nur die F2-Analyse signifikant wurde, sich in den F1-Analysen aber nur Tendenzen aufzeigten.

ein Kontext, in dem die 1. Person Singular als Gewaltverbrecher beschrieben wird. Des Weiteren stellt Bergen und Wheeler (2005) eine Studie vor, in der auch für Sätze ein ACE gefunden wurde, in denen der Agens in der 3. Person Singular steht. Dies könnte dadurch begünstigt worden sein, dass außer dem Agens kein weiterer belebter bzw. menschlicher Referent im Satz genannt wurde.

**Simulation und Anaphernaufflösung** Die zweite Studie, die in der Freiburger Forschungsgruppe durchgeführt wurde<sup>24</sup>, führt uns zurück zum Thema des vorangegangenen Kapitels: Sie beschäftigt sich mit der Referenzzuweisung bei der Anaphernaufflösung in der Online-Sprachverarbeitung und untersucht hierfür weitere Constraints, die aus der Embodied-Cognition-Perspektive resultieren. Ihr liegt eine Reihe von Annahmen für die Verarbeitung von Anaphern zugrunde:

- Anaphern sind in gerichteten motorischen Handlungen verankert – genauer gesagt in deiktischen Zeigegesten. Hinweise auf eine Begründung von Anaphern in deiktisch verwendeten Demonstrativen finden sich in Ansätzen der diachronen Theorie der Grammatikalisierung (vgl. z.B. Diessel, 2006): Demonstrativa sind sprachübergreifend als historische Quelle für die Entwicklung grammatischer Marker wie definite Artikel, Relativ- und Personalpronomen oder Konjunktionen belegt. Dabei teilen sie sich die funktionale Eigenschaft, gemeinsame Aufmerksamkeit auf einen Diskursreferenten herzustellen (in der realen Welt oder übertragen in einem Text).
- Beim Sprachverstehen wird ein mentales Situationsmodell in einem mentalen Situationsraum konstruiert (z.B. Zwaan, 2004).
- Die inkrementelle Positionierung der im Satz benannten Elemente in diesem mentalen Situationsraum erfolgt nach einem Präferenzschema. Das bedeutet, wenn keine spezifischen linguistischen Angaben von Ordnungsrelationen zur Verfügung stehen, wird eine stereotypische Ordnung aufgebaut. Hinweise darauf fanden sich beispielsweise in Eyetracking-Studien, in denen sich für die abgebildeten Referenten unerwartete Positionierungseffekte zeigten (vgl. Hemforth et al., 2009; Konieczny & Weldle, 2009).
- Die Verarbeitung von Anaphern schließt die Simulation einer entsprechenden Zeigegeste in diesem mentalen Situationsraum ein. Diese Zeigegeste ist gerichtet und verweist deiktisch auf das Element in diesem Situationsraum, das als Antezedens der Anapher identifiziert wurde.
- Eine sprachlich ausgelöste Simulation einer Handlung kann mit einer tatsächlich auszuführenden Handlung interagieren und zu Kompatibilitätseffekten (z.B. Glenberg & Kaschak, 2002) oder Interferenz-Effekten (z.B. Bergen, 2005) führen.

Die einzige Annahme, die nicht auf vorhergehenden Studien begründet ist, betrifft die durch Anaphern induzierte Simulation von Zeigegesten im mentalen Raum. Nach dieser Hypothese wird erwartet, dass die simulierten Zeigegesten im Sinne des ACE mit korrespondierenden oder entgegengesetzten Handlungen interagieren, die tatsächlich ausgeführt

---

<sup>24</sup>In der Zwischenzeit wurde die Studie zu einem mit Drittmitteln finanzierten Projekt (Anschubfinanzierung durch die Universität Freiburg) ausgeweitet, innerhalb dessen mehrere Experimente mit weiteren Variationen und zusätzlichen Kontrollmaßnahmen durchgeführt werden.

werden. Je nachdem, wo sich die angenommene Person im mentalen Raum befindet, verweist die Zeigegeste in eine entsprechend andere Richtung. Für die präferierte Anordnung von benannten Referenten im mentalen Raum wurde aus den Nebeneffekten früherer Studien angenommen, dass diese Referenten inkrementell von links nach rechts positioniert werden. Je nachdem, wann eine Person im Satz genannt wurde, verweist die simulierte Zeigegeste also nach rechts oder nach links auf einer horizontalen Achse.<sup>25</sup> Wird in der Folge die tatsächliche Ausführung einer nach links oder nach rechts gerichteten Zeigegeste erforderlich gemacht, so sollte sich ein Effekt der Kompatibilität zwischen simulierter und auszuführender Bewegungsrichtung zeigen. Diesen horizontalen deiktischen ACE bei der Verarbeitung von Anaphern haben wir als Pointing Compatibility Effect (PCE, Konieczny & Weldle, 2009; Konieczny, Weldle, Haser et al., 2010; Konieczny, Haser et al., 2010) bezeichnet.

In der Studie wurden Sätze verwendet, die zwei Referenten einführen, auf die in einem folgenden Nebensatz über ein Personalpronomen verwiesen wird (vgl. Bsp. 32). Die beiden möglichen Antezedenzen unterscheiden sich im Genus. Je nach Benennungsreihenfolge werden sie gemäß der Hypothese links oder rechts im mentalen Raum positioniert. Durch das Genus des Personalpronomens wird disambiguiert, welcher der genannten Referenten – das Satz-Subjekt oder das Satz-Objekt – sein korrektes Antezedens ist. Im Nebensatz wurde zudem eine Handlung beschrieben, die in einer stereotypischen Relation zu einem der beiden benannten Referenten steht: In unserem Beispielsatz etwa stellt die Briefträgerin einen Brief zu (Bsp. 32b und 32d) und der Müllmann leert die Mülltonne (Bsp. 32a und 32c). In vier Bedingungen wurde die Positionierung (rechts oder links) und das Genus von Antezedenzen und Pronomen (Maskulinum oder Femininum) variiert. Für die Aufgabe, die die VTn im Experiment durchführen mussten, wurden für alle Sätze weitere Versionen erstellt, die eine nicht plausible bzw. auffällig untypische Situation implizieren, indem sie die in den Nebensätzen beschriebene Handlung dem jeweils anderen Referenten zuweisen (im Beispiel: Der Müllmann stellt den Brief zu und die Briefträgerin leert die Mülltonne).

- (32)
- a. Der Müllmann sah die Briefträgerin, als er die Mülltonnen leerte.  
*Bewegungsrichtung links kompatibel*
  - b. Der Müllmann sah die Briefträgerin, als sie den Brief zustellte.  
*Bewegungsrichtung rechts kompatibel*
  - c. Die Briefträgerin sah den Müllmann, als er die Mülltonnen leerte.  
*Bewegungsrichtung rechts kompatibel*
  - d. Die Briefträgerin sah den Müllmann, als sie den Brief zustellte.  
*Bewegungsrichtung links kompatibel*

Die Aufgabe der VTn war es, die Sätze zu lesen, und anschließend per Tastendruck die Plausibilität bzw. Typikalität der Sätze zu beurteilen. Die Sätze wurden mit der Methode

---

<sup>25</sup>Einschließlich der Annahme des Zeigens im mentalen Raum sind diese Annahmen natürlich spekulativ: Die präferierte Anordnung könnte durch die Leserichtung induziert oder anderweitig kulturell bedingt sein (vgl. ähnliche Aspekte bei Santiago et al., 2007). Darüber hinaus legt eine Positionierung von links nach rechts nur die Reihenfolge der Referenten im mentalen Raum fest – dies muss aber nicht zwingend bedeuten, dass ein später genannter Referent in Relation zu meiner Zeigeposition tatsächlich rechts von mir steht. Hinweise auf eine solche Positionierung finden sich in der Gebärdensprache in der kanonischen Verortung sukzessive eingeführter Personen im Gebärdenraum.

RSVP im Moving Window-Paradigma präsentiert. Dazu mussten die VTN auf einer Buttonbox eine zentral gelegene Taste gedrückt halten. Nach Abschluss des Satzes mussten sie die Beurteilung durch das Drücken einer links oder rechts auf der Buttonbox gelegenen Taste durchführen.<sup>26</sup> Die Belegung dieser Tasten wurde nach der Hälfte des Experiments getauscht.

Für die Daten wurde ein lineares gemischtes Modell gefittet. Es finden sich zwei zentrale Effekte: Erstens wurde die Beurteilung generell schneller durchgeführt, wenn mit dem Personalpronomen auf das Satzsubjekt verwiesen wurde. Dies entspricht den Erwartungen für das Deutsche, für das eine allgemeine Präferenz für die Anbindung an das Satzsubjekt besteht (vgl. Hemforth et al., 2009). Zweitens zeigte sich ein Kompatibilitätseffekt gemäß unseren Hypothesen. Wenn die Handlungsrichtung für die Beurteilung kompatibel zur Zeigerichtung der simulierten deiktischen Geste ist, erfolgt die Beurteilung schneller als in den Bedingungen mit inkompatibler Handlungsrichtung.

Trotz der positiven Befunde, die unsere Hypothesen unterstützen, lässt diese Studie eine große Anzahl Fragen offen. Aus diesen lassen sich interessante Fragestellungen für Folgeuntersuchungen ableiten. So stellt sich etwa die Frage des Einsatzes und der Persistenz des Effekts: Findet sich für ihn eine Abhängigkeit von der motorischen Planung der auszuführenden Handlung (vgl. Borreggine & Kaschak, 2006)? Wie lange bleibt die Aktivierung der motorischen Simulation erhalten und wann setzt sie ein? Und ist es nicht erstaunlich, dass wir noch nach Off-Set des Satzes einen Effekt finden, obwohl die durch die Anapher induzierte motorische Simulation bereits vier Wörter zurückliegt und am Ende des Satzes zudem eine andere Handlung (die Handlung des Nebensatzes) benannt wird, die ja ebenfalls eine motorische Simulation bedingen könnte? Die größte Schwäche der Studie liegt wohl in den Annahmen zur Konstruktion und Gliederung des mentalen Raums: Wie sind die Präferenzen bei der stereotypischen Konstruktion tatsächlich belegt, wie stark sind diese ausgeprägt und in welchem Maße sind sie kulturell bedingt? Dies beinhaltet auch Bezüge zu Fragen linguistischer Relativität bei der Wahrnehmung und Verarbeitung von Raum, für die der Einfluss unterschiedlicher Orientierungen und räumlicher Perspektiven diskutiert wird (vgl. etwa Levinson, 2003; Boroditsky, 2003; Papafragou, 2008). Allerdings ist hinsichtlich der kulturellen Bedingtheit auch denkbar, dass die gefundenen Kompatibilitätseffekte nicht aus der Interaktion mit Simulationen in einem mentalen Raum resultieren, sondern schlicht Effekte der Kompatibilität mit der Leserichtung sind.<sup>27</sup> Des Weiteren kann hinterfragt werden, ob die Effekte tatsächlich einen anaphorischen Verweis benötigen, der eine entsprechende Handlungssimulation auslöst, oder ob nicht allein der Verweis auf eine Person (etwa durch die stereotypische Handlungsbelegung im Nebensatz) für einen entsprechenden Effekt genügt: Statt einer simulierten Zeigegeste könnte eine

---

<sup>26</sup>Diese Handlung sollte einer Zeigegeste nahekommen und musste mit dem Zeigefinger ausgeführt werden. Inzwischen läuft ein weiteres Experiment, bei dem diese Aufgabe auf einem Touchscreen erfolgt. Diese Realisierung der Aufgabe erfordert, dass eine tatsächliche Zeigegeste ausgeführt werden muss. Sie ist damit näher an der Annahme einer simulierten Zeigegeste, mit der die Aufgabe interagieren soll.

<sup>27</sup>Im Deutschen ist die Leserichtung von links nach rechts vorgegeben – in Entsprechung zur angenommenen Präferenz beim Aufbau des mentalen Raums. Um auszuschließen, dass die gefundenen Effekte schlicht Effekte der Kompatibilität zur Leserichtung sind, werden derzeit im Psycholinguistischen Forschungslabor Freiburg weitere Experimente durchgeführt, die diesen Einfluss minimieren. Natürlich könnte die Leserichtung auch eine Ursache dafür sein, dass mentale Räume präferiert von links nach rechts gegliedert werden – doch dann sollten sich die Effekte auch bei Kontrolle des Einflusses der Leserichtung (etwa durch auditive Präsentation der Stimuli) zeigen.

simulierte Blicksteuerung bzw. gerichtete Aufmerksamkeitlenkung im mentalen Raum für den Effekt verantwortlich sein. Nicht zuletzt wäre noch ein interessanter Aspekt, ob umgekehrt linguistische Verarbeitungspräferenzen – etwa die Subjektpräferenz bei der Auflösung von Personalpronomen im Deutschen – durch Stimuli räumlicher Bewegung moduliert werden können. Können also durch Wahrnehmung und Handlung bedingte neuronale Aktivierungen Einfluss auf linguistische Situationssimulationen nehmen? Auch hierfür wurden Hinweise gefunden (vgl. Zwaan & Taylor, 2006; Kaschak et al., 2005; Zwaan & Yaxley, 2003).

## 9. Embodied Cognition und Konnektionismus

Im zweiten Kapitel zum Thema Embodied Cognition werden wir uns näher mit den Beziehungen zwischen diesem Forschungsfeld und dem Konnektionismus beschäftigen. Einerseits ist das Verhältnis zwischen diesen beiden Perspektiven auf Kognition nicht gänzlich frei von Differenzen. Andererseits gibt es doch auffällige Gemeinsamkeiten und überlappende Grundlagen sowie gemeinsame Beziehungen zu weiteren verwandten Forschungsansätzen. Betrachtet man Erwerb, Repräsentation und Verarbeitungsprozesse, findet man zwangsläufig verwandte Konzepte und Termini. Um einige Schlagwörter zu nennen: Emergenz kognitiver Strukturen und Prozesse, induktiv-assoziative Lernmechanismen, Analogizität (Inkorporierung struktureller Bezüge des Gegenstands) und Prototypizität der Kategorisierungsprozesse, Verarbeitung als Constraint-Satisfaction-Prozess, Verwischung der Unterscheidung von Verarbeitung und Repräsentation.

Wir werden uns zunächst dem übergreifenden Feld der Modellierungsansätze für Language Grounding und Embodied Language Processing nähern. In der Forschung zu Künstlicher Intelligenz (KI) finden sich frühe Ansätze zur Integration von visueller und sprachlicher Verarbeitung. Weitere Arbeiten sind stärker auf die Struktur und den Erwerb von Sprache bezogen, oder überprüfen mittels autonomer Agenten und Roboter Theorien zur Entstehung von Sprache. Die meisten dieser Ansätze sind vor allem deswegen für uns interessant, weil sie grundlegende Fragen der Modellierung dieser Gegenstände aufwerfen. Einen Großteil an den Modellierungsansätzen nehmen die konnektionistischen Modelle ein (vgl. Cangelosi, 2005). Bevor wir uns mit diesen beschäftigen, wollen wir zuerst allgemein Beziehungen zwischen Konnektionismus und Embodied Cognition beleuchten: Dies betrifft gleichermaßen gemeinsame Grundlagen und Konzepte sowie Abweichungen und Unterschiede der beiden Rahmentheorien. Dabei werden wir klären, inwiefern konnektionistische Modelle einen geeigneten Rahmen für die Modellierung von Aspekten der Embodied Cognition bieten können und welchen Nutzen wir uns aus der Entwicklung einer gemeinsamen Perspektive erhoffen. Unseres Erachtens ist einerseits für konnektionistische Modelle der Sprachverarbeitung eine theoretische Schärfung der Konzepte Referenzialität, Systematizität und Bedeutung zu erwarten. Andererseits erwarten wir eine konzeptionelle und terminologische Schärfung für die Embodied Cognition sowie die Möglichkeit, kognitive Modelle für diesen Forschungsansatz zu entwickeln, um die mitunter vagen Theorien weiter zu spezifizieren. Im Anschluss geben wir einen Überblick über Ansätze, die eine (eliminativ) konnektionistische Perspektive mit der Embodied Cognition in Verbindung bringen (u.a. Plunkett et al., 1992; Coventry et al., 2005; S. L. Frank et al., 2009). Viele Modelle der Satzverarbeitung wurden zur Embodied Cognition in Beziehung gesetzt. Es bleibt allerdings zu hinterfragen, welche Kriterien entsprechende Modelle erfüllen müssen, um den Anforderungen der Embodiment-Ansätze gerecht zu werden.

## 9.1. Modellierungen von Embodied Language Processing

Beim folgenden Überblick zu Modellierungsansätzen für die Embodied Cognition beschränken wir uns weitgehend auf Modelle der Sprachverankerung (Language Grounding), der Integration von sprachlicher und visueller Verarbeitung und des Embodied Language Processing. Hierfür wurden sehr unterschiedliche Herangehensweisen und Modellierungsmethoden vorgeschlagen. Die Ansätze reichen von symbolisch orientierten Produktionssystemen über hybride Ansätze zu Kommunikationssystemen für Roboter und simulierte autonome Agenten bis hin zu konnektionistischen Netzwerk-Modellen. Wir werden uns in diesem ersten Unterkapitel mit Ansätzen beschäftigen, die keinen primär eliminativ-konnektionistischen Ansatz wählen. Hier finden sich bereits früh Ansätze aus der Informatik bzw. der KI-Forschung. Diese versuchen vor allem, die Leistungen der maschinellen Sprachverarbeitung zu verbessern, indem sie das Symbol-Grounding-Problem für Sprache über die Integration von linguistischer und visueller Verarbeitung überwinden. In diesem Zusammenhang werden verschiedene Ansätze von konnektionistischen und hybriden bis zu symbolischen Modellen diskutiert. Verschiedene hybride Architekturen und der sogenannte strukturierte Konnektionismus finden sich in der Forschung zur situierten und gegroundeten Sprachverarbeitung, etwa in den Arbeiten der L<sub>0</sub>-Gruppe (vgl. Feldman et al., 1996) und in der Neural Theory of Language (NTL, Feldman & Narayanan, 2004), in deren Rahmen auch Ansätze zur Modellierung der Embodied Construction Grammar (ECG) gehören. Ansätze zur Modellierung des Ursprungs und der Entwicklung von Sprache und ihrer Verankerung in der kommunikativen Interaktion stammen aus der Robotik (z.B. Steels, 2003) und aus der Simulation autonomer Agenten (z.B. Cangelosi & Riga, 2006).

### 9.1.1. Integration linguistischer und visueller Verarbeitung

Frühe Ansätze für die Modellierung situiert oder an der Wahrnehmung orientierter Sprachverarbeitung finden sich in der KI-Forschung (vgl. etwa das Special Issue des *Artificial Intelligence Review*, McKevitt, 1996). Ein wichtiges Ziel dieser Ansätze ist es, die Leistungen der maschinellen Sprachverarbeitung zu verbessern. Eine mögliche Herangehensweise hierfür ist, sich an der menschlichen Sprachverarbeitung zu orientieren, die im Vergleich zu künstlichen Sprachverarbeitungs- und Übersetzungssystemen sehr gut funktioniert. Ein als wichtig erachteter Unterschied ist dabei, dass Menschen Sprache situiert und damit in der Wahrnehmung und der Interaktion mit der Umwelt verankert erwerben. Als unabdingbare Voraussetzung für die Überwindung grundlegender Probleme der künstlichen Systeme wurde deshalb die Berücksichtigung des Symbol-Grounding-Problems angenommen. Auf den Arbeiten von Searle (1980) und Harnad (1990) basierend wird das darin involvierte Problem des Zirkelschlusses diskutiert (z.B. Marconi, 1996a; McKevitt & Guo, 1996; Jackson & Sharkey, 1996). Marconi (1996b) stellt dabei fest, dass die kognitive Verarbeitung der Bedeutung eines Wortes mehr sein muss als die Übertragung von Informationen von einem System (Sprache) in ein anderes System (sozusagen die Sprache des Gehirns) im Sinne einer Übersetzung, denn wir verfügen bei der Aktivierung der Bedeutung über Handlungs- und Anwendungswissen in der realen Welt. Dem stellt er die Hypothese der referenziellen Inkompetenz künstlicher Systeme der natürlichen

Sprachverarbeitung gegenüber (Marconi, 1996a). Erst, wenn ein System Sprache auf die wahrgenommene Welt beziehen kann, verfügt es über semantische Kompetenz und kann Sprache im eigentlichen Sinne verstehen. Deshalb muss die Grundlage eines echten Systems natürlicher Sprachverarbeitung zumindest die Integration (künstlicher) Perzeption beinhalten und dabei echte Objekterkennung leisten können.<sup>1</sup>

Eine solche Lösung des Grounding-Problems über die Integration linguistischer und visueller Verarbeitung wurde verschiedentlich vorgeschlagen. In diesem Zusammenhang wurden verschiedene Ansätze von konnektionistischen über hybride bis zu symbolisch-algorithmischen Modellen diskutiert. Dabei stellt sich aber die Frage, wie ein solches integratives System aussehen und wie in seiner Implementierung die situational-perzeptionelle Verankerung von Bedeutung geleistet werden soll. Darüber hinaus wäre aber noch grundsätzlicher zu fragen, welche Voraussetzungen ein Körper erfüllen muss, um Ausgangspunkt eines verkörperlichten Systems sein zu können, wann ein System beanspruchen kann, tatsächlich verkörperlicht zu sein, und was es für ein Symbol qualitativ bedeutet, gegroundet zu sein (vgl. Chrisley, 2003). Genügt es, ein System mit einem Roboterarm und einer Videokamera auszustatten, damit es die Bedingungen einer Embodied Cognition erfüllt? Gegen einen solchen Roboteransatz wurden Einwände im Sinne des Chinese-Room-Arguments hervorgebracht (vgl. Searle, 1980): Wird ein Roboter mit einem Videosystem ausgestattet, kann er den diskret-abstrakten internen Symbolen letztlich nur sinnfreie Perzepte ohne semantischen Gehalt zuweisen. Marconi (1996a) beleuchtet diesen Einwand aus verschiedenen Blickwinkeln und stellt ihm die Position entgegen, dass wir auch unser menschliches Verstehen nur aus einer solchen Perspektive erklären können – denn schließlich können wir nur aus unserer subjektiven Innenperspektive konstatieren, dass wir einer Sache tatsächlich semantischen Gehalt zuweisen.

**Die Frage der Implementierung** Für die Implementierung eines in der visuellen Verarbeitung gegroundeten Systems natürlicher Sprachverarbeitung wurden verschiedene Probleme identifiziert (u.a. Marconi, 1996a; Kasabov, Postma & van den Herik, 2000):

- Das künstliche visuelle System darf kein Rekordersystem im Sinne eines Bilderspeichers sein, sondern muss Schematisierung und Konzeptualisierung leisten (vgl. auch Barsalou, 1999).
- Darüber hinaus darf das System auch nicht auf das Erkennen von Objektformen reduziert sein, sondern muss auch funktionale Zusammenhänge erkennen. Um dies zu leisten, genügt eine reine Top-down-Objekterkennung (im Sinne eines Listenabgleichs) nicht mehr: Das System muss in der Lage sein, Top-down- und Bottom-up-Prozesse zu integrieren.<sup>2</sup> Auf den technischen Stand seiner Zeit bezogen schließt

---

<sup>1</sup>In der Regel wird in den besprochenen Modellen das Grounding von sprachlichen Äußerungen in Bezug auf den Gegenstand ihrer Referenz angenommen. Der Bezug zwischen der visuellen und der akustischen Wahrnehmung von Sprache – also nicht von Sprache auf einen visuellen Referenzgegenstand, sondern zwischen den unterschiedlichen Erscheinungsvormen von Sprache als akustischer und optischer Reiz durch Gestik und Lippenbewegungen – wird zum Beispiel von Thompson und Ogden (1995) oder Robert-Ribes, Schwartz und Escudier (1995) besprochen.

<sup>2</sup>Marconi (1996a) erläutert dies an dem Satz *Es liegt etwas/ein Spielzeug auf dem Tisch*. Sehr viele Objekte mit unterschiedlicher Form können als Spielzeug dienen; und mit dem Wort *etwas* ist das Objekt schließlich gänzlich unterspezifiziert. Um den Satz dennoch als wahr erkennen zu können, muss zunächst

Marconi (1996a), dass kein künstliches System aktuell auch nur annähernd die referenzielle Kompetenz des Menschen nachbilden kann.

- Die Vielschichtigkeit der visuellen und linguistischen Verarbeitung betreffend fragt Marconi (1996a) für die technische Ebene: Wo sollen die Verbindungsstellen für referenzielle Prozeduren im künstlichen visuellen System verortet werden und wo interagieren die verschiedenen Ebenen des linguistischen Systems?<sup>3</sup>
- Welche Systeme eignen sich hinsichtlich ihrer repräsentationalen und komputationalen Kapazität als Grundlage der Implementierung?
- In Bezug auf Repräsentation und Integration: Soll eine gemeinsame, über verschiedene Modalitäten hinweg abstrahierende bzw. vermittelnde Repräsentation angenommen werden? Oder existieren viele modalitätsspezifische Repräsentationen, zwischen denen vermittelnde Interfaces angenommen werden müssten?

Für diese Fragen der Implementierung, dem zugrundegelegten System und der Realisierung der Integration wurden verschiedene Lösungsvorschläge gemacht, die wir hier aber nur kurz nennen wollen. Neben symbolischen wurden konnektionistische und hybride Ansätze in ihrer Eignung diskutiert (vgl. Jackson & Sharkey, 1996). Ein Überblick über symbolisch orientierte, in der klassischen KI verankerte Ansätze zur Integration linguistischer und visueller Informationen ist zu finden bei Srihari (1995). In einem symbolischen Ansatz, der mitunter etwas oberflächlich und naiv anmutet, schlagen McKevitt und Guo (1996) ein integriertes (multimodales) mentales Lexikon vor, in dem sprachliche Formen auch hinsichtlich räumlicher und visueller Strukturen definiert sind. Ebenfalls auf Symbolen und Algorithmen basiert das in Lisp implementierte System von Meini und Paternoster (1996), in dem lokative Ausdrücke in einem integrativen Verifikationssystem mit visuellem Input abgeglichen werden. In eher formal-symbolischen Ansätzen entwickeln Siskind (1995) und Cohn, Gooday, Bennett und Gotts (1995) Logikkalküle für räumliche Konzepte als Grundlage einer Sprach-Bild-Integration. Ein hybrides Modell schlägt Grumbach (1996) vor. In diesem erfolgt das Grounding von diskret-abstrakten Symbolen anhand eines subsymbolischen assoziativ-sensomotorischen Substrats. Ein rein assoziativer Ansatz des Grounding wird von Sales, Evans und Aleksander (1996) vorgestellt. Hier lernen Neural State Machines (einer konnektionistischen Architektur, die auf gewichtsfreien künstlichen Neuronen basiert) die Objektbenennung auf Grundlage der Assoziation statischer Graustufen-Bilder mit Wörtern.

**Die Frage der Repräsentation** Die Frage der modalitätsübergreifenden vs. modalitätsspezifischen Repräsentation sowie die Frage der Unterschiede vs. Zusammenhänge der

---

bottom-up ein Erkennensprozess stattfinden, um anschließend seine (funktionale) Kategorienzugehörigkeit zu überprüfen.

<sup>3</sup>Kasabov et al. (2000) formulieren weitere zu bestimmende Spezifikationen in einem allgemeineren Ansatz für die Integration von visueller und akustischer Verarbeitung. Beide Modalitäten sind hierarchisch und in mehreren Ebenen aufgebaut: (a) Auf welche Weise und auf welchen Ebenen der Modalitäten werden Informationen integriert, (b) auf welche Weise wird Zeit und unterschiedliche Temporalität realisiert und wie erfolgt die Synchronisation des Informationsflusses der verschiedenen Quellen, (c) wie soll das adaptive Lernen für die verschiedenen Modalitäten erfolgen und wie wird neues Wissen aus den einzelnen Modalitäten integriert? Diese Fragen sind für alle Modelle multimodaler Verarbeitung zu berücksichtigen.

Verarbeitungsmechanismen verschiedener Modalitäten wird zum Beispiel von Partridge (1995), Schank und Fano (1995) und Robert-Ribes et al. (1995) diskutiert. In diesen Ansätzen wird das Verstehen in der Regel als modalitätsunabhängiger Prozess konzipiert, wofür aber verschiedene Argumentationen herangezogen und verschiedene Realisierungen dieser Unabhängigkeit vorgeschlagen werden. Dessen zugrundeliegende unifizierte Struktur wird zum Beispiel durch einen singulären perzeptuellen Mechanismus (Partridge, 1995) oder durch einen Fusionsmechanismus zur Strukturierung der Informationen innerhalb eines gemeinsamen Repräsentationsraums (Robert-Ribes et al., 1995) gewährleistet. Schank und Fano (1995) hingegen entfernen sich von dieser input-zentrierten Sichtweise und betrachten die Informationen aus verschiedenen Modalitäten als Aspekte einer übergeordneten, zu lösenden Aufgabe, die modalitätsunabhängig im Zentrum des Verstehensprozesses steht.

### 9.1.2. Erwerb und Verarbeitung sensomotorisch verankerter Sprache

Ein groß angelegtes Projekt, das neben der Modellierung von visueller und sprachlicher Verarbeitung auch die Modellierung des Spracherwerbs und speziell des Bedeutungserwerbs anstrebt, ist das L<sub>0</sub>-Projekt (vgl. Feldman et al., 1996). Als fundamentale Aufgabe des Systems wurde formuliert, dass es aus Beispielen von Bildern gepaart mit korrespondierenden Beschreibungen (d.h. wahren Statements über die Bilder in einer arbiträren natürlichen Miniatur-Sprache) die Sprache gut genug lernen muss, um für neue Sätze der Sprache bestimmen zu können, ob sie ein wahres Statement für ein begleitendes Bild abgeben. Kann das System diese Sprach-Bild-Verifikation leisten, wird ihm erfolgreiches Verstehen zugesprochen. Die methodische Vorgabe ist dabei ein möglichst theoriefreies Vorgehen bei der Implementierung zum Beispiel hinsichtlich Grammatik-Formalisierungen und einzubauenden Verarbeitungsmechanismen – das Modell soll die gesamte Abbildungsleistung selbst erbringen. Um sich diesen Zielen zu nähern, wurde das Projekt in mehrere Teilprojekte zu spezielleren Fragestellungen aufgeteilt, in denen sehr unterschiedliche Systeme zum Einsatz kamen.

Von besonderem Interesse sind in diesem Zusammenhang die Arbeiten von Regier (Regier, 1995, 1996). In einer konnektionistischen Modellierung simuliert er den Erwerb der sprachlichen Kategorisierung statischer und dynamischer räumlicher Relationen (z.B. *auf, neben, durch*). Dabei soll das Modell die Entwicklung von Präpositionen verschiedener Sprachen und ihren gemeinsamen Kern der räumlichen Kategorisierung abbilden. Die menschliche Kapazität zum Erwerb räumlicher Semantik wird dabei auf grundlegende visuelle Wahrnehmungseigenschaften zurückgeführt – diese stellen die gemeinsame Basis für sprachliche Universalien. Die Modellarchitektur schließt eine Anzahl struktureller Mechanismen mit ein, die durch neurobiologische und psychophysische Evidenzen motiviert sind. Diese Mechanismen stellen den Kern der Constraints dar, auf denen die Universalien der Raumkategorisierung beruhen. Das Modell ist in der Tradition des adaptiven strukturierten Konnektionismus implementiert (vgl. z.B. Feldman, Fianty, Goddard & Lynne, 1988). In diesem wird eine Verbindung von Prinzipien neuronaler Netzwerke und klassischer KI (d.h. Inferenzen und Repräsentationen anhand von Algorithmen und Datenstrukturen) innerhalb eines komputationalen Systems realisiert. Damit ist das Modell von Regier (1995) architektonisch eher der Klasse hybrider Modelle zuzurechnen als dem eliminativen Konnektionismus. Dennoch ist es ein früherer Ansatz, Embodied Language Processing

in konnektionistischen Modellen zu simulieren.<sup>4</sup> Zwei weitere komputationale Ansätze beziehen sich stärker auf die Aspekte des motorischen Grounding sprachlicher Semantik (vgl. Feldman et al., 1996; Bailey, 1997; Narayanan, 1997; Feldman & Narayanan, 2004).<sup>5</sup>

Diese Ansätze können der Neural Theory of Language (NTL, Feldman & Narayanan, 2004) subsumiert werden, die als aktuelle Fortsetzung des L<sub>0</sub>-Projekts zu verstehen ist. In der NTL wird noch stärker der Bezug zur Embodied Cognition hervorgehoben. Ausgangspunkt für die Rahmentheorie ist das Gehirn als System spezifischer neuronaler Strukturen, und die Fragen, wie diese die Natur des Denkens und der Sprache formen sowie in welcher Form Denken und Sprache in Beziehung zu anderen neuronalen Systemen stehen. Damit rückt die neurofunktionale Realisierung der Verankerung von Bedeutung in der sensomotorischen Erfahrung in den Vordergrund. Ein zentrales Ziel innerhalb dieser Rahmentheorie ist die Entwicklung von Modellen, die mit bekannten Daten und beschränkenden biologischen, psychologischen, linguistischen und komputationalen Bedingungen konsistent sind. Prominent werden dabei vor allem strukturierte konnektionistische Modelle vertreten (vgl. Chang, Feldman & Narayanan, 2005). In den Rahmen der NTL ist auch die ECG (Bergen & Chang, 2003) einzuordnen, die als komputational umsetzbare Formalisierung für Konstruktionen im Sinne des Embodied Language Processing zu verstehen ist. Anhand der komputationalen Modelle soll die Rolle der körperlich verankerten Repräsentationen und der durch die Simulation beim Sprachverstehen ermöglichten Inferenzen untersucht werden. Dafür wurde ein Parser entwickelt (Bryant, 2004b, 2004a), der linguistische Strukturen anhand eines Lernmodells auf Grundlage der ECG-Formalismen und eines Unifikationsmechanismus analysiert.

### 9.1.3. Weitere Ansätze zum Sprach- und Bedeutungserwerb

Eine ganze Reihe weiterer Modellierungsansätze beschäftigen sich mit Fragen des Sprach- und Bedeutungserwerbs, und untersuchen dabei, welche Rolle es für den Erwerb spielt, dass er situiert und modalitätsübergreifend erfolgt. Ein zentrales Problem betrifft dabei die Frage des Vokabelerwerbs bzw. Bedeutungserwerbs: Beim Erstspracherwerb steht das Kind der Aufgabe gegenüber, eine Flut von Informationen zu strukturieren, dabei sprachliche Formen zu identifizieren und diese schließlich auf ihre Bedeutung abzubilden. Die Aufgabe der Identifikation der relevanten Einheiten wird auch als Bootstrapping-Problem bezeichnet. Nachdem sprachliche Äußerungen von allgemeinen Geräuschen differenziert, in einzelne Elemente segmentiert und als bedeutungsvoll erkannt wurden, bleibt als Hürde die Zuweisung von Bedeutung. Diese Aufgabe ist alles andere als trivial, denn es besteht keine einfache Eins-zu-eins-Korrespondenz zwischen Wörtern und Bedeutungen: Zum

---

<sup>4</sup>Einen neueren Ansatz auf dieser Grundlage beschreiben Regier und Carlson (2001) im Attentional Vector-Sum-Modell (AVS) und einer darauf bezogenen empirischen Studie zum Grounding von raumbezogener Sprache in der Wahrnehmung.

<sup>5</sup>Bailey (1997) stellen ein Modell vor, das die Semantik von Handlungsverben und ihren Erwerb als Grounding der sprachlichen Ausdrücke in bereits entwickelten motorischen Schemata simuliert. Auf dieser Verankerung von Verbsemantik in Handlungsschemata aufbauend implementiert Narayanan (1997) ein Modell des Metaphernverstehens, das im Sinne der Conceptual Metaphors (Lakoff & Johnson, 1980) motorisch-räumliche Handlungsschemata als Quellen-Domäne für abstrakte Konzepte oder Metaphern auffasst. citeARoy2005 stellt einen ähnlichen auf Schemata basierenden Ansatz vor, in dem neben Handlungen weitere Kategorien wie Objekte, Eigenschaften und Relationen erfasst werden.

einen kann sich eine Äußerung auf ganz verschiedene Aspekte einer Situation beziehen, zum Beispiel auf die gesamte Situation, ein Ausschnitt einer Situation, einzelne Ereignisse oder Objekte einer Situation oder die persönliche Einstellung des Sprechers zu einer Situation (referenzielle Indeterminiertheit/Unsicherheit, vgl. Quine, 1960). Zum anderen bestehen sprachliche Äußerungen zumeist nicht nur aus einem, sondern aus mehreren Wörtern, die zudem auf ganz verschiedene Aspekte von Referenzsituationen verweisen, zum Beispiel Nomen auf Objekte und Personen, Verben auf Zustände und Ereignisse, Präpositionen auf räumliche Relationen von Objekten und Handlungsbezügen. Weitere Hürden kommen hinzu, zum Beispiel das Rauschen im Input sowie die Synonymie und die Homonymie sprachlicher Ausdrücke.

Das lernende Kind steht also einer kombinatorischen Explosion möglicher Zuweisungen zwischen mehreren Wörtern und mehreren Situationsaspekten gegenüber: Es muss eine komplexe Abbildung multipler sprachlicher Formen auf multiple, raumzeitlich korrespondierende Situationen, Situationsaspekte oder Gebrauchsmöglichkeiten leisten, und dafür aus einem Möglichkeitsspektrum von Viele-zu-viele-Korrespondenzen eine Eins-zu-eins-Korrespondenz etablieren. In Bezug auf diese Problematik relativieren Rohde und Plaut (1999) ihre Kritik an der „Importance of starting small“-Hypothese von Elman (1993). Während Limitierungen der Stimuli oder der kognitiven Kapazität für den Erwerb syntaktischer Relationen irrelevant sein mögen (vgl. 3.1.3.1), können sie aber dennoch förderlich für das Sprachverstehen und den Bedeutungserwerb sein. Denn Verstehen erfordert über den Erwerb grammatischer Strukturen hinaus die komplexe Assoziation von Bedeutung (bzw. von Situationen und Ereignissen der Umwelt, die diese Bedeutung konstituieren) mit sprachlichen Oberflächenformen. Dieser Prozess dürfte durch einfache bzw. vereinfachte Äußerungen mit einer beschränkten und konkreten Bedeutung wesentlich erleichtert werden.

**Etablierung von Referenzbeziehungen** Davis (2002) bespricht in einer Übersicht mehrere Lösungsansätze, in denen versucht wird zu erklären, wie das spracherwerbende Kind den vom Sprecher intendierten Referenzgegenstand einer Äußerung erkennt (vgl. auch F. Kaplan, Oudeyer & Bergen, 2007). Herangezogen wurden hierzu unter anderem Gestaltprinzipien, Salienz, gemeinsame gerichtete Aufmerksamkeit und Aufmerksamkeitslenkung durch den Sprecher. Als Lösungsansatz für Mehrwortäußerungen wird in verschiedenen Ansätzen neben der Integration multimodaler Informationen (etwa im Sinne eines wechselseitigen Bootstrapping-Mechanismus) das situationsübergreifende Lernen (cross-situational learning, z.B. Gleitman, 1990) herangezogen. Nach diesem wird beim Bedeutungserwerb – etwa anhand situationaler Schnittmengenbildung oder Prototypisierung – aus mehrfachen Kookkurrenzen von Wörtern und verschiedenen Situationen ein eingegrenzter Bedeutungsgehalt einer Äußerung ermittelt. Es wurden verschiedene komputationale Modelle vorgeschlagen, die diese Aspekte der Integration multimodaler Informationen zur Lösung des Korrespondenzproblems beim Bedeutungserwerb thematisieren.<sup>6</sup>

Siskind (1996) stellt einen Algorithmus auf Grundlage einer logischen Formalisierung der Aufgabe der Abbildung von Wort zu Bedeutung vor. Dieser leistet situationsübergrei-

---

<sup>6</sup>Mit diesen wird auch die Beschränkung der Spracherwerbsdebatte auf die Angeborenheit und Lernbarkeit sprachlicher Strukturen und Kompetenzen überwunden, die zum Beispiel von F. Kaplan et al. (2007) bemängelt wird.

fendes Lernen bzw. situationsübergreifende Inferenzen bei referenzieller Indeterminiertheit im Sinne einer Schnittmengenbildung über propositionalen Kookkurrenzen einer sprachlichen Äußerung. Für Mehrwort-Äußerungen wird dabei das Prinzip der Kontrastierung zwischen den verschiedenen Wörtern umgesetzt. Damit konkretisiert Siskind das schwammige Konzept des situationsübergreifenden Lernens und ermöglicht die Evaluation der Effektivität dieses Ansatzes. In der Modellierung operiert der Algorithmus in Form von Regeln auf propositional strukturierten Situationen und korrespondierenden sprachlichen Äußerungen und entwickelt dabei ein mentales Lexikon für Form-Bedeutungs-Paare.

Das CELL-Modell (cross-channel early lexical learning, Roy, 1999) beschreibt den Prozess der Abbildung von Wort zu Bedeutung als Suche nach Strukturen über verschiedene Modalitäten. Dabei verwendet es Rohdaten eines Mikrophons und einer Videokamera als Input ohne jegliche externe Aufbereitung der Daten. Das erworbene semantische Wissen des Modells entspricht also perzeptuell gegroundeten Kategorien. Die Rohdaten werden in atomare prototypische Bestandteile zerlegt. Dabei können alle Bestandteile der Äußerung potenziell auf alle Bestandteile des Kontexts verweisen. Durch informationstheoretische Filter- und Cluster-Prozesse werden aus diesen Paaren lexikalische Einheiten generiert. Die Abbildung von Form auf Bedeutung erfolgt als situationsübergreifender Inferenzprozess über Gedächtnissysteme: Das Kurzzeitgedächtnis speichert anhand eines Kookkurrenzfilters ausgewählte, zusammen auftretende Äußerungs-Kontext-Ereignisse. Anhand eines Rekurrenz-Filters werden situationsübergreifend mehrfach auftretende Äußerungs-Kontext-Ereignisse identifiziert und als lexikalische Hypothesen gespeichert. Lexikalischen Hypothesen mit hohem Informationsgehalt gehen in das Langzeitgedächtnis über.

Im Ansatz von Yu und Ballard (2003) erfolgt der Worterwerb über die Interaktion des Modells mit einem menschlichen Benutzer (vgl. auch Yu & Ballard, 2004). Im Zentrum steht die Idee, dass die Verankerung sprachlicher Symbole in der Wahrnehmung und die Abbildung zwischen Form und Bedeutung stark an Inferenzen über die referenziellen Intentionen des Sprechers gebunden sind, die sich in seiner sichtbaren gerichteten Aufmerksamkeit (z.B. über Blick-, Kopf- und Handbewegungen) äußern. In einem hybriden multimodalen Modell wird diese Idee umgesetzt. Die Aufgabe des Systems ist die Generierung eines Lexikons für Objekte aus kurzen Sätzen für visuelle Szenen. Die multimodalen Rohdaten erhält es dabei von einem menschlichen Informanten, der mit einer Kamera, einem Eyetracker, einem Mikrophon und Positionssensoren an Kopf und Händen ausgerüstet ist, und die von ihm gesehenen Objekte beschreibt. Bei der Verarbeitung werden die nichtsprachlichen Kontextinformationen genutzt, um die Segmentierung und Erkennung von Wörtern zu unterstützen. Zudem werden Informationen über die Gerichtetheit der Aufmerksamkeit dazu genutzt, multimodale raumzeitliche Korrelationen zu identifizieren und darüber die Bedeutung lexikalischer Einheiten zu binden. Die Form-Bedeutungs-Abbildung wird über Cluster- und Expectation-Maximization-Algorithmen geleistet, die durch die modalitätsübergreifenden Constraints informiert werden. Umgekehrt zeigt sich durch die Integration multimodaler Informationen auch eine Verbesserung der Objektidentifizierung und -kategorisierung durch die linguistischen Informationen.

In einem Ansatz zum statistischen Lernen präsentieren Yu und Smith (2006) Ergebnisse aus Experimenten und Simulationen zum statistischen situationsübergreifenden Lernen. Die Modellierung basiert auf formalen Repräsentationen und einem Gedächtnissystem mit zwei repetitiven Algorithmen. Hypothetische Form-Bedeutungs-Paare werden im Ge-

dächtnis gespeichert. Diesen wird auf der Grundlage bestätigender und konfligierender Informationen aus anderen Beispielen einen Konfidenzwert zugewiesen. Neue Paare werden auf der Grundlage eines Filter-Algorithmus aufgenommen oder abgelehnt, der die Konsistenz des Paares mit den bestehenden Paaren prüft. Dabei muss über alle Paare eine strikte Eins-zu-eins-Korrespondenz erhalten bleiben. Für den Filter-Algorithmus werden nur die Paare als Bedingungen berücksichtigt, deren Konfidenzwert eine bestimmte Schwelle überschreitet.<sup>7</sup>

### 9.1.4. Sprachgenese in Robotern und autonomen Agenten

Weitere Ansätze für Language Grounding und Embodied Language Processing stammen aus der Robotik und aus Simulationen autonomer Agenten und sind dem Forschungszweig Artificial Life zuzuordnen. In diesen wird vor allem die Genese von Kommunikation und von sprachlichen Strukturen thematisiert. Ausgangspunkt sind Gruppen von Agenten, die in einer Umwelt situiert sind und über einen Körper bzw. über sensomotorische Interaktionsmöglichkeiten verfügen. Die Wahrnehmungen, Handlungen und Interaktionen können simuliert sein oder in der tatsächlichen physischen Welt realisiert werden. Wir besprechen kurz die beiden genannten Ansätze. Eine erweiterte Diskussion zum Thema mit einer Übersicht zu bestehenden Modellen bietet Cangelosi (2005).

Im Roboter-Ansatz wird versucht, die Evolution von Sprache als Effekt situierter sozialer Interaktion von Individuen zu erklären, die über eine bestimmte körperliche Ausstattung für Wahrnehmung, Handlung und kognitive Verarbeitung in der realen Welt verfügen (für einen Überblick siehe Steels, 2003; Cangelosi, 2005). Im Zentrum der Untersuchung steht dabei, wie sich aus der situierten Interaktion strukturierte (im Sinne von konventionalisierten) Kommunikationsmuster entwickeln. Hierzu werden künstliche Agenten mit einer Peripherie für die Wahrnehmung ihrer Umwelt (z.B. visuelle und auditive Verarbeitung, haptische Sensorik) und physisch-motorische Handlungen (z.B. ein Roboterarm, ein beweglicher Kopf, eine Entsprechung zum Vokaltrakt) ausgestattet. Für die Verarbeitung von Wahrnehmungseindrücken, die Handlungsplanung und für Kategorisierungsprozesse können verschiedene Systeme zugrundegelegt werden. Auch künstliche neuronale Netzwerke kommen als zentrale Berechnungseinheit zum Einsatz. In Experimenten mit vorgegebener Aufgabe interagieren die Roboter-Agenten mit anderen Robotern oder mit Menschen. Aus solchen Interaktionssituationen und -aufgaben sollen sich Kommunikationssysteme entwickeln, die mit natürlichen Kommunikationssystemen vergleichbar sind und eventuell Aufschlüsse über diese geben. Sehr bekannt geworden sind die „Talking Heads“-Experimente von z.B. Steels, Kaplan, McIntyre und van Looveren (2002). In diesen führen zwei mit Videosystem ausgestattete Roboter-Agenten Sprachspiele durch, in denen sie wechselseitig erkennen müssen, auf welches Objekt in der Umwelt der jeweils andere sprachlich referiert. Ein Wahrnehmungssystem leistet die Distinktion und prototypische Kategorisierung der Objekte in der Umwelt, ein assoziatives Gedächtnissystem ermöglicht die Speicherung von ausgehandelten Form-Bedeutungs-Paaren. Durch gegenseitige Hemmungen konkurrierender Belegungen für Form-Bedeutungs-Paare und über

---

<sup>7</sup>In einem weiteren Modell zeigt Yu (2006), dass die Form-Bedeutungs-Abbildung auf der Grundlage von Mehrwortäußerungen durch die Berücksichtigung syntaktischer Informationen im Sinne des syntaktischen Bootstrappings weiter gestützt werden kann.

verstärkende Feedback-Schleifen werden lexikalische Elemente etabliert – es entwickelt sich selbstorganisierend ein konventionalisiertes Lexikon. Mit diesem Experiment wird der Austausch über die Welt in einer Gruppe von Agenten und die Herstellung geteilter Aufmerksamkeit als Grundlage menschlicher Kommunikation vorgeschlagen. Ähnliche Experimente wurden zum Beispiel auch herangezogen, um die Entwicklung grammatischer Strukturen als Effekt von Interaktion und Konsens zwischen Agenten zu simulieren. Die Entwicklung von Kommunikation sowie von geordneter Umweltkategorisierung findet in diesen Systemen ohne vorhergehende Spezifizierungen oder sonstige menschliche Interventionen statt. Einen ähnlichen Ansatz verfolgen Dominey und Boucher (2005) mit einer hybriden Architektur mit konnektionistischem Kernstück (vgl. Kap. 9.3.2.2).

In Simulationen virtueller autonomer Agenten werden Gruppen adaptiver autonom handelnder Teilprogramme einer virtuellen Umwelt ausgesetzt. Diese Agenten können mittels virtueller Sensorik und Motorik mit der Umwelt und mit den anderen Agenten interagieren. In einer solchen Simulation untersuchen Cangelosi und Harnad (2000) den evolutionären Vorteil von Symbolic Theft-Mechanismen. Symbolic Theft beschreibt die Entwicklung abstrakter und ungegründeter komplexer Konzepte durch die Kombination von einfachen sensomotorisch geordneten Symbolen (vgl. auch Cangelosi et al., 2002). Ziel der Agentenpopulation ist es, in der virtuellen Umwelt zu überleben, wofür die Agenten Nahrungssuche betreiben müssen. Dazu müssen sie erfahrungsbasierte Kategorienbildung (sensomotorisch geordnet) leisten, um zwischen essbaren und giftigen Pilzen unterscheiden zu können (trial-and-error-Lernen). Alternativ zur Fähigkeit, eigene Erfahrungen für die Kategorisierung zu nutzen, können die Agenten Informationen austauschen und ihre Entscheidungen aufgrund dieser Informationen treffen – somit trägt auch die Entwicklung einer konventionalisierten Benennung zur Kategorienbildung für die Pilze bei (Benennungslernen). Anhand genetischer Algorithmen werden zudem Selektion und Reproduktion implementiert: Nur die hinsichtlich der Nahrungssammlung am besten angepassten Agenten pflanzen sich fort. In vergleichenden Simulationen zeigen sich deutliche evolutionäre Vorteile für Agentenpopulationen, die ausschließlich die Kommunikationsstrategie nutzen.

In einem weiteren Ansatz, der die Symbolic Theft-Hypothese stützen soll, erwerben virtuelle epigenetische Roboter Wissen über Basishandlungen – einfache Interaktionen mit Objekten – auf der Grundlage der Imitation beobachteten Verhaltens. Zudem erwerben sie sprachliche Benennungen, die in diesen sensomotorischen Basishandlungen geordnet sind (Cangelosi & Riga, 2006). Der Fokus verlagert sich hier vom Grounding in der Perzeption zum Grounding in der motorischen Interaktion. Das motorische und linguistische Verhalten der Roboter wird über künstliche neuronale Netzwerke gesteuert. Neue behaviorale Fähigkeiten höherer Ordnung werden anhand von Übertragungsmechanismen durch sprachliche Interaktionen mit menschlichen Versuchsleitern im Sinne einer sprachlichen Handlungsanleitung aus den geordneten Basishandlungen konstruiert und auch benannt. Die Simulationen demonstrieren eine klare Konzeptionierung des Übertragungsmechanismus, der für die angenommenen Prozesse der Verankerung ungegründeter in geordneten sprachlichen Symbolen erforderlich ist.

## 9.2. Eine gemeinsame Perspektive

Im vorangegangenen Unterkapitel haben wir Modellierungsansätze besprochen, die Sprachverstehen aus der Perspektive der Embodied Cognition simulieren oder an diese angelehnt sind. Neben diesen Ansätzen gibt es eine ganze Reihe an Modellierungen mit konnektionistischen Netzwerken, die zu Embodiment und Language Grounding in Beziehung gesetzt wurden. Im nächsten Unterkapitel gehen wir zuerst einen Schritt zurück, um aus einer übergeordneten Perspektive zu beleuchten, ob eine gemeinsame Perspektive aus Konnektionismus und Embodied Cognition möglich und sinnvoll ist. Denn es gilt zu berücksichtigen, dass sich diese Forschungsansätze in verschiedener Hinsicht nicht decken. Allerdings teilen sie sich eine Vielzahl gemeinsamer Konzepte wie etwa Constraint Satisfaction, Erfahrungsbasiertheit und Emergenz, was eine gewisse Kompatibilität andeutet. Unseres Erachtens bieten sich Konnektionismus und Embodied Cognition eine ganze Reihe sinnvoller theoretischer und methodischer Ergänzungen, die dabei helfen können, gegenseitig die Profile ihrer Positionen zu kognitiven Prozessen und Strukturen zu schärfen. In diesem Sinne versuchen wir im Folgenden, einige Fragen zu beantworten: Erfüllen konnektionistische Modelle überhaupt die Voraussetzungen, um Aspekte des Embodiment zu simulieren? Sind solche Modelle sinnvoll und nützlich – können mit ihnen also theoretische Ansätze konkretisiert oder empirische Vorhersagen getroffen werden? Was nützt die Embodiment-Perspektive der eliminativ konnektionistischen Auffassung von Kognition und Sprachverstehen und inwiefern stärkt oder unterstützt sie diese Auffassung? Dazu besprechen wir zuerst kritische Sichtweisen aus der Sicht der Embodied Cognition und fassen kurz einige zentrale gemeinsame Konzepte zusammen. Abschließend diskutieren wir den Nutzen, die Chancen und mögliche Synergien einer gemeinsamen Perspektive.

### 9.2.1. Gemeinsamkeiten und Differenzen

In vielen Studien der beiden Rahmentheorien finden sich immer wieder Verweise auf Gemeinsamkeiten und Unterschiede zwischen Konnektionismus und Embodied Cognition.<sup>8</sup> So finden sich bereits früh Hinweise auf das Potenzial konnektionistischer Ansätze, Probleme des Language Grounding zu überwinden (Harnad, 1990; Plunkett et al., 1992). Und viele aktuelle Modelle stellen einen Bezug zu Ideen der Embodied Cognition her (u.a. Altmann & Mirkovic, 2009; Mayberry et al., 2009; S. L. Frank et al., 2009). Auch von Seiten der Embodiment-Forschung werden Zusammenhänge beleuchtet, wenn auch nicht immer ganz optimistisch (u.a. Glenberg, 1997; Ellis, 1998; Barsalou, 1999, 2008). Trotz der gegenseitigen Annäherungen in grundlegenden Fragen zur Kognition bleibt zu prüfen, inwiefern überschneidend verwendete Termini tatsächlich das Gleiche meinen. Wir besprechen (a) Repräsentation, (b) Erwerb, (c) Entwicklung von Eigenschaften, (d) Verarbeitung, (e) neuronale Orientierung und (f) materiellen Bindung.

<sup>8</sup>Auch wenn man nicht wirklich von Rahmentheorien im Sinne eines übergreifenden und das Forschungsfeld vereinheitlichenden Ansatzes sprechen kann, werden wir in der folgenden Diskussion der Einfachheit halber Konnektionismus und Embodied Cognition als Rahmentheorien bezeichnen. Wir versuchen, uns bei den Zuschreibungen von Grundannahmen an den größten gemeinsamen Nennern innerhalb der Rahmentheorien zu halten. Wir legen dabei die zentralen Charakterisierungen zugrunde, die wir in den Kapiteln 2.2 und 8.1 entwickelt haben. Zu einzelnen Zusammenhängen und Unterschieden finden sich Zusammenfassungen unter anderem in Ellis (1998); MacWhinney (2001); Rohrer (2006).

**Gemeinsamkeiten** Die Konzeption von Repräsentation ist in beiden Rahmentheorien stark von Analogizität geprägt. Die internen bzw. mentalen Repräsentationen und Strukturen erhalten die Zusammenhangs- und Ähnlichkeitsstruktur der Referenzobjekte und derer Zusammenhänge in der repräsentierten Welt. In der Embodied Cognition ist dies dadurch gegeben, dass die Repräsentationen Derivate der Wahrnehmung von der und Interaktion mit der Welt sind und nicht in abstrakte diskrete Symbole übertragen werden. Im Konnektionismus wird dies durch subsymbolische verteilte Repräsentationen gewährleistet, die übertragen in einen hochdimensionalen Zustandsraum Repräsentationen in räumliche Verortungen und Zusammenhänge in Distanzen und Trajektorien abbilden. In beiden Rahmentheorien sind Repräsentationen dynamisch und erfahrungsbasiert. Die Analogizität bezieht sich demzufolge auch auf die Sensitivität der Systeme gegenüber Frequenzen, Verteilungen und die Zeitlichkeit von Beispielen, die sich auf die Verarbeitungseigenschaften niederschlagen. Die Beispielsbasiertheit des Erwerbs im Sinne situationalen Lernens ermöglicht einen gemeinsamen Ausgangspunkt für das Grounding Problem. Der Erwerbprozess wird für beide Rahmentheorien als induktiv sowie durch bereits entwickelte kategoriale Strukturen gestützt beschrieben. Die sich entwickelnden kategorialen Strukturen im System sind beiderseits stark durch Effekte der Prototypikalität, Kontextualität und Gradualität bestimmt. Beim Erwerb werden verschiedene aufeinander bezogene Muster miteinander assoziiert, er erfolgt über induktiv-assoziative Lernmechanismen.<sup>9</sup>

Der Erwerb und die Verarbeitung wird innerhalb beider Rahmentheorien immer wieder als Constraint-Satisfaction-Prozess beschrieben. Eine Interpretation der Welt erfolgt als Aushandlung verschiedener Faktoren der Umwelt (raumzeitlich, physisch, sozial) und des Systems (Erfahrung, aktueller Zustand/Kontext, Lern- und Wahrnehmungsmechanismen). Die internen Repräsentationsstrukturen und die Eigenschaften des Systems entwickeln sich in Abhängigkeit der Erfahrungen und Aushandlung – dies wird als Emergenz bezeichnet.<sup>10</sup> Bei Erwerb und Verarbeitung werden niedrigere (z.B. Wahrnehmung) und höhere (z.B. Inferenz) kognitive Prozesse durch dieselben kognitiven Prinzipien und Mechanismen bestimmt: Die Domänen Wahrnehmung und Kognition lassen sich in der Embodied Cognition nicht trennen, da höhere kognitive Prozesse auf der Basis der neuronalen Derivate der Wahrnehmung operieren. Im Konnektionismus ist die Einheitlichkeit der Grundprozesse und -prinzipien durch die Reduzierung auf einfache und identische funktionale Einheiten implizit gegeben. Sowohl im Konnektionismus als auch in der Embodied Cognition entsprechen Verstehens- und Denkprozesse der constraint-basierten Aktivierung erfahrungsabhängiger Strukturen. Dies entspricht im Konnektionismus der Transformation von Aktivierungsmustern bei der Propagierung von Aktivierung durch das Netzwerk und in der Embodied Cognition der Simulation oder dem Re-Enactment von schematischen modalen

---

<sup>9</sup>Hier ist allerdings anzumerken, dass Glenberg (1997) Assoziation als erklärendes Konzept gänzlich ablehnt. Dem Assoziationismus als Konzept der Beschreibungsebene im Sinne konditionaler Wahrscheinlichkeit von Zuständen stellt er das Meshing gegenüber, das durch körperliche Constraints und Affordanzen den Akt der Assoziation erst bedeutungsvoll macht.

<sup>10</sup>Als Emergenz wird die spontane, nichtüberwachte Strukturbildung in einem System durch die Interaktion seiner Bestandteile verstanden. Ein komplexes System kann demzufolge nur anhand eines dynamischen Erklärungsansatzes erfolgen, der dessen Emergenz erfasst, denn das entstehende System kann nicht aus den einzelnen Bestandteilen erklärt werden, sondern nur aus den Interaktionen, aus denen das System hervorgeht. Zu Erläuterungen des Emergenz-Begriffs und seinen Beziehungen zum Konnektionismus und zur Embodied Cognition vgl. Elman et al. (1996); Ellis (1998); Stephan (1999); MacWhinney (2001); O'Grady (2008); McClelland et al. (2010).

Erfahrungsderivaten (z.B. Barsalou et al., 2003) oder dem Meshing-Prozess (Glenberg, 1997).

**Grenzen des Konnektionismus** Wir sind über Repräsentation, Erwerb und Verarbeitung bis zu den Aspekten der neuronalen Realisierung und damit zur materiellen Grundlage der Rahmentheorien gekommen – und hier zeigen sich die Inkompatibilitäten, die sich wieder zurück bis zu den Repräsentationen auswirken. In der Embodied Cognition ist der Körper mit seinen Eigenschaften zentrale Quelle der anzunehmenden Constraints und Affordanzen. Er gibt durch seine Modalitäten und seine Funktionsweise Informationsstrukturen und -kanäle sowie Interaktionsmöglichkeiten vor. Zwar teilt der Konnektionismus einige zentrale Konzepte und Eigenschaften der Repräsentation und Verarbeitung. Zudem teilt er sich aus einer möglichen eliminativ materialistischen Sicht mit dem Embodiment das gemeinsame ‚Feindbild‘ klassisch-funktionalistischer Ansätze. Doch hinsichtlich der tatsächlichen Ausprägung der materiellen Eigenschaften bleibt er neutral. Damit kann er zentrale Aspekte der Embodied Cognition nicht inhärent erfassen, denn der Kern der Bedeutungshaftigkeit entsteht laut Embodiment-Theorien nicht aus reinen Assoziationen, sondern aus Assoziationen in Bezug zu den Eigenschaften und Möglichkeiten des Körpers. Diese Grenze zwischen den Rahmentheorien wird durch Glenberg (1997, S. 8) sehr deutlich formuliert:

Connectionist accounts of semantic or meaningful information are based on conceiving of meaning as activation of a limited number of features, at least at the input layer. Unfortunately, most theorists fail to specify what the features are, and they fail to specify how those features might be learned or changed as a consequence of development. In the system that I am proposing, initial coding is not featural, but analog, in terms of patterns of possible action. Furthermore, as one learns more about the interactive capabilities of one's body, objects and actions can be imbued with new meaning: what I can do with that object now. Another important difference concerns the nature of constraints. In standard connectionist accounts, constraints are [...] extrinsic [...]. That is, a particular constraint represents statistical, or joint-occurrence, information, not a necessary feature of the operation of the system. Thus, a connectionist system would be equally happy to learn that a Coke bottle can be used as a chair or as a weapon. In an embodied system, constraints arise because of analog coding of projectable features and their implications for human action. [...] [T]hese constraints are intrinsic to the operation of the system. For example, how we think about a Coke bottle is constrained not just by particular experiences with Coke bottles, but by the actual shape and heft of the bottle, too. Thus, an embodied system would have little difficulty understanding how a Coke bottle could be a weapon, but it would balk at learning that it could be used as a chair.

Auf diesen Umstand und weitere Inkompatibilitäten weist auch Elman (1999a) hin, wenn er fundamentale Unzulänglichkeiten konnektionistischer Modelle beschreibt: (a) Sie sind passive reaktive Modelle, die kein selbsttätiges zielgerichtetes Verhalten entwickeln; (b) sie sind (sofern sie nicht Bestandteil autonomer Agenten-Systeme sind) asoziale Modelle, die Kognition als individuelles Phänomen erfassen; und (c) sie sind entkörperte Modelle, und können deshalb zentrale Constraints unseres Erlebens und Denkens nicht beschreiben. Als relativ neutrale materielle Plattform verfügen konnektionistische Netzwerke eben nicht im Vorhinein über die Spezifika, die für die Embodied Cognition entscheidend werden. Der

Konnektionismus bietet lediglich eine Umsetzungsmöglichkeit für Theorien der Embodied Cognition und nicht bereits die Lösung. Welche Form die relevanten Constraints annehmen müssen, bleibt das Forschungs- und Entwicklungsziel der Modellierer. Hierzu noch einmal (Glenberg, 1997, S. 8):

These comments should not be taken to mean that an embodied system cannot be simulated using connectionism. In fact, it may well be that connectionism will be the surest route to formalizing these ideas. Nonetheless, it will have to be a connectionism that differs from the sorts currently in use.

### 9.2.2. Chancen und Nutzen einer gemeinsamen Perspektive

Wie wir gesehen haben, teilen sich die beiden Rahmentheorien viele gemeinsame Konzepte und ähnliche Ausgangspunkte. Diese Gemeinsamkeiten führen auch zu vielen Überschneidungspunkten der Perspektiven auf die kognitiven Leistungen des Sprachverstehens und -erwerbs. Andererseits gibt es aber auch Unstimmigkeiten zwischen den Rahmentheorien, etwa das Fehlen physischer und psychophysischer Constraints als erklärender Faktor im Konnektionismus. Diese Lücken müssen beim Versuch der Modellierung entweder überbrückt oder aber aus dem Erklärungsspektrum ausgeschlossen werden. Es stellt sich die Frage, ob ein solcher reduzierter Ansatz sinnvoll ist bzw. ob die Herstellung überbrückender Vorbedingungen für einen vollständigen Modellierungsansatz gelingen kann. Außerdem ist noch die Sinn- und Nutzenfrage zu stellen: Was nützt uns eine wechselseitig ergänzende Perspektive, und inwiefern erweitert dies theoretische Ansätze des Konnektionismus und der Embodied Cognition? Wie wir im Folgenden begründen werden, ist aus einer wechselseitig ergänzenden Perspektive eine nützliche Symbiose im Sinne einer gegenseitigen konzeptuellen Erweiterung und theoretischen Schärfung zu erwarten.

**Bedarf komputationaler Modelle** Wir können zunächst feststellen, dass von Seiten der Embodiment-Forschung verschiedentlich der Bedarf an komputational implementierten Modellen geäußert wurde. Speziell Barsalou (2008) bemängelt, dass die Orientierung an empirisch-experimentellen Herangehensweisen zu intuitiven Grundannahmen der Embodied Cognition nur der erste Schritt ist. Nachdem für viele Annahmen positive Befunde erbracht wurden, charakterisiert er es als nächste Stufe der Theoriebildung und folglich als wichtige Herausforderung der Embodied Cognition, wohlspezifizierte Theorien anhand konsistenter formaler und komputationaler Modelle zu entwickeln. Regier (in Glenberg, 1997, S. 34) stellt dazu fest, dass bereits die Umsetzung und die Feststellung der dazu erforderlichen Spezifizierungen wichtige Schritte sind:

If after all the trouble of implementation we are left with the conclusion that it is the general principles we began with that count, why bother? The answer is that the process of modeling can lead to a clarification of these general principles. Prior to jumping to implementation, a careful examination at the computational and algorithmic levels (Marr 1982) would determine the content of the data structures required and the nature of the operations running over them. And this, rather than the implementation itself, could help in making Glenberg's attractive but still largely intuitively grasped theoretical notions more concrete. That would definitely be worthwhile.

Es fragt sich allerdings: Wie kann Embodied Cognition überhaupt modelliert werden? Wie kann ein Ansatz, der Kognition unter anderem anti-funktionalistisch und anti-repräsentationalistisch (z.B. Markman & Dietrich, 2000) konzeptualisiert, so weit spezifiziert, formalisiert und operationalisiert werden, dass er sich sinnvoll in einem komputationalen kognitiven Modell implementieren lässt? Eine formale Theorie setzt nun mal voraus, dass sich in einem System innerhalb axiomatischer Bedingungen durch Symbole und Operatoren wohldefinierte Zustände bestimmen lassen. Wie sich Bedingungen des menschlichen Körpers und dessen Zustände sinnvoll in ein formales System überführen lassen sollen, ist für uns nicht nachvollziehbar. In der Tat gibt es bereits Ansätze wie etwa die ECG (Bergen & Chang, 2003) im größeren Rahmen der NLT (Feldman & Narayanan, 2004). Die ECG wurde zuvor in den Kapiteln 8.2.4 und 9.1.2 besprochen und genau in dieser Hinsicht problematisiert. Noch deutlicher kritisiert S. L. Frank (2007) in einem Review einer zentralen Monographie zur NTL, dass bei den hierin entwickelten Versuchen, die Embodied Cognition in ein formales System zu überführen, über eine Hintertür gerade wieder die Prinzipien abstrakt-amodaler symbolischer Ansätze eingeführt werden, die die Embodied Cognition überwinden wollte. So bleibt zu fragen, ob eine formale Theorie der Embodied Cognition sinnvoll oder nicht eher kontraproduktiv ist.

**Eignung konnektionistischer Modelle** Demgegenüber bietet der Konnektionismus die Möglichkeit, Modelle zu entwickeln, die keine Formalisierungen im Sinne der Bestimmung diskreter mentaler Repräsentationen, Strukturen, Mechanismen und algorithmischer Operationen benötigen. Spezifischen Formalisierungen wie etwa in GOFAI-Modellen steht hier eher ein System grundlegender Prinzipien gegenüber, anhand dessen Aspekte der Embodied Cognition simuliert werden können. Selbst in der einschränkenden Beurteilung von McCloskey (1991) entwickeln wir damit zwar keine Realisierungen von Theorien, erhalten aber doch die Möglichkeit einer Methodik zur Theorieentwicklung (vgl. Kap. 2.1). Selbst wenn wir damit nur Modelle im Sinne von Untersuchungsobjekten erhalten, so wäre dies bereits ein lohnenswerter Schritt, über den Theorien der Embodied Cognition überprüft werden können.

Allerdings glauben wir, dass konnektionistische Modelle noch mehr zu bieten haben. Denn erstens bleiben auch in diesen Modellen Aspekte zu spezifizieren – dies betrifft neben architektonisch gewählten Kategorisierungsmechanismen vor allem die Bestimmung der Formate der Stimuli sowie die Aufgaben, die mit der Transformation dieser Input- und Output-Strukturen geleistet werden müssen. Dies können relevante Gesichtspunkte sein, die eine Spezifizierung der Embodied Cognition weiterführen. Zweitens lassen sich anhand der hier wirkenden Prinzipien die vagen Annahmen von Embodiment-Theorien über kognitive Prozesse spezifizieren – etwa Constraint-Satisfaction-Mechanismen, Effekte von Frequenz und Erfahrung, die Realisierung der Integration verschiedener Informationen. Im Sinne der Emergenz-Theorie werden in der Embodied Cognition komplexe Wechselwirkungen zwischen Körper, Umwelt und Erfahrungen angenommen. Effekte solcher Wechselwirkungen lassen sich nur schwer intuitiv abschätzen, können aber mithilfe des Werkzeugs Konnektionismus modelliert werden. Drittens können anhand der Performanz und ihrer Beschränkungen falsche Annahmen der intuitiv formulierten Theorien identifiziert oder benötigte Erweiterungen für einen voll-funktionalen Ansatz vorgeschlagen werden. Damit bietet das theoretische Instrumentarium des Konnektionismus eine geeignete Plattform

zur Implementierung und Simulation von Embodied Cognition und Embodied Language Processing.

Ein Beispiel eines solchen Ansatzes, Embodiment-Theorien konnektionistisch zu spezifizieren, geben Joyce, Richards, Cangelosi und Coverntry (2003) mit dem Connectionist Perceptual Symbol System Network (CPSSN). Mit diesem Modell schlagen die Autoren SRNs als mögliche Grundlage zur konnektionistischen Implementierung perzeptueller Symbolsysteme im Sinne der Rahmentheorie von Barsalou (1999) vor, und versuchen auf dieser Grundlage zu erfassen, welcher Natur entsprechende Repräsentationen im Sinne der Embodied Cognition sein müssen. Joyce und Kollegen erfassen in ihrer Modellierung einige prinzipielle Aspekte, die für die Verarbeitung visueller Informationen für perzeptuelle Symbolsysteme angenommen werden.<sup>11</sup> Dabei diskutieren sie Fragen der Kompositionalität der Repräsentation: Sowohl die vermuteten Anforderungen an ein entsprechendes System als auch die Modellierung legen die Annahme verteilter überlagerter Repräsentationen nahe. Damit wird direkt die Position Barsalous kritisiert, der in seinen strukturellen Beschreibungen perzeptueller Symbolsysteme klassische syntaktisch-kompositionelle Repräsentationen vermutet. Eine ähnliche Position bezieht auch Jackson und Sharkey (1996), wenn er feststellt, dass die mit dem Konnektionismus implementierte räumliche Abbildung von verteilten Repräsentationen in einem Zustandsraum für die Konzeptionen der Embodied Cognition sehr vorteilhaft ist, da deren nicht-konkatentative Kompositionalität die Erfassung von neuen und ad hoc-Kombinationen von Objekten, Zuständen und Eigenschaften ermöglicht. Zudem ermöglicht der Konnektionismus die Interaktion von Repräsentationen und Aufgaben verschiedener Domänen. Dies ist ein relevanter Aspekt der Entwicklung von Repräsentationen in der Embodied Cognition, der in anderen Modellierungsansätzen nicht mit dieser Einfachheit erreicht werden kann (Reilly, 1995).

**Nutzen für konnektionistische Ansätze** Während der Nutzen konnektionistischer Modellierungen für Theorien der Embodied Cognition auf der Hand liegen, muss umgekehrt die Frage gestellt werden, ob die Übernahme von Annahmen der Embodied Cognition auch von Nutzen für die Entwicklung einer konnektionistischen Perspektive auf kognitive Verarbeitung und speziell Sprachverstehen ist. Unserer Ansicht nach stehen hier drei Gesichtspunkte im Vordergrund: Die Unterstützung der Systematizitätsfrage, die Konkretisierung angenommener Repräsentationen und die Erweiterung der Perspektive konnektionistischer Modellierungen.

Während es auf den ersten Blick nicht so scheinen mag, gibt es zwischen den Kriterien, die für eine Untersuchung konnektionistischer Systematizität (vgl. Kap. 3.1.4) und für eine Modellierung von Sprachverarbeitung im Sinne der Embodied Cognition herangezogen werden müssen, direkte Bezüge. In Reaktion auf die Kritik von Fodor und Pylyshyn (1988), konnektionistische Modelle könnten nur dann Systematizität vorweisen,

---

<sup>11</sup>Auch wenn häufig die visuelle Domäne herangezogen wird, so ist das Universum möglicher sensomotorischer Domänen in konnektionistischen Netzwerken nicht auf die visuelle Verarbeitung beschränkt. In einem Ansatz, der dem Modell von Joyce et al. (2003) in einigen Punkten ähnelt, stellen Botvinick und Plaut (2004) die Modellierung von normalen und eingeschränkten sequenziellen Routine-Handlungen ohne die Verwendung klassischer Schema-Hierarchien vor, und simulieren damit die Kontrolle motorischer Handlungssequenzen.

wenn sie ein klassisch-symbolisches System implementieren, problematisieren S. L. Frank et al. (2009) die Umsetzung des Satzverstehens in verschiedenen konnektionistischen Ansätzen. Speziell richten sie sich dabei gegen die Verwendung von Stimuli, mit denen klassisch-semantische Repräsentationen mit expliziter Vorgabe semantischer Strukturen in das Modell eingeführt werden. Da solche Modelle im Verdacht stehen, anhand dieser Strukturen selbst wieder nur Implementierungen klassisch-symbolischer Systeme zu sein, kann mit diesen kein valides Gegenbeispiel gegen die Systematizitätskritik am Konnektionismus etabliert werden. Möchte man also die Systematizität entsprechender Modelle belegen, muss sichergestellt sein, dass die verwendeten Stimuli nicht durch symbolische Informationsstrukturen die Arbeitsweise des Modells beeinflussen. Eine ähnliche, aber eher pragmatische Kritik äußert Waskan (2001) an gängigen Verfahren der Enkodierung von Stimuli in konnektionistischen Ansätzen semantischer Verarbeitung. Dabei erläutert er seinen Verdacht, dass explizite semantische Repräsentationen inhärent durch die Intentionen des Modellierers beeinflusst sind. Eine Möglichkeit der Umgehung besteht in der Verwendung von kookkurrenz-statistischen Repräsentationen, etwa der LSA. Allerdings ist dieses Repräsentationsformat wiederum zu flach (Waskan, 2001, S. 288): „In particular, representations of semantic relatedness do not suffice to capture information about the many spatial, causal and functional relationships that the objects described in a given sentence bear to one another.“

Wie könnten Stimuli enkodiert sein, die die Verwendung explizit-symbolischer Strukturen und die Beeinflussung durch den Modellierer vermeiden? Eine mögliche Lösung bietet die Orientierung an den Vorgaben der Embodied Cognition. So wählen etwa S. L. Frank et al. (2009) ein Repräsentationsformat, das jegliche explizite semantische Information eliminiert, und nur noch hinsichtlich zeitlicher Korrelationen (z.B. Gleichzeitigkeit oder Aufeinanderfolgen) bestimmt ist. Diesem Format schreiben die Autoren zu, im Sinne der Forderungen von (Barsalou, 1999) die Analogizität der erfahrenen Welt zu erhalten. Wir werden auf diese Diskussion nochmals in Kapitel 9.3.2 eingehen. Auch Waskan (2001) schlägt eine Lösung zur Befreiung von der Last des Enkodierens vor, die mit Ansätzen der Embodied Cognition und des Language Grounding vereinbar sind (S. 288):

Ultimately, we should like to find that the state of the art is such that perceptual mechanisms are able to ground, directly or indirectly, the contents of the semantic representations harboured by models of comprehension, thus effectively freeing modellers from the responsibility of generating semantic representations. Until such a time, however, there may be no alternative but to rely, at least in part, upon human semantic intuitions.

Werden ausschließlich die Aspekte der Umwelt enkodiert, die wir auch selbst wahrnehmen können, umgeht man auch das Problem, das sich mit der Verwendung von Target-Vektoren einstellt. Verwendet man für die Verarbeitung im Netzwerk Zielvorgaben für Transformationen, die das Netzwerk leisten soll, so müssen diese theoretisch begründet sein. Denn schließlich sind sie ein Bestandteil der Erklärung der Netzwerk-Performanz. Eine selbsterklärende Rechtfertigung ergibt sich, wenn man Enkodierungen und Transformationsaufgaben verwendet, die in der natürlich gegebenen Situation in der Umwelt vorliegen.

Abseits der Systematizitätsfrage könnte man die Embodied Cognition als Chance sehen, die konnektionistische Perspektive auf bislang übergangene Aspekte des Spracherwerbs und der Sprachverarbeitung zu erweitern und die hierfür relevanten Faktoren zu kon-

kretisieren. Dies betrifft zum Beispiel die von (Glenberg, 1997) kritisierten theoretisch unterspezifizierten semantischen Merkmale bei der Stimulus-Enkodierung. Die Enkodierung solcher Merkmale als Eigenschaften und Zustände sensomotorischer Verarbeitung könnte Ansätze eröffnen, wie semantische Merkmale überhaupt erst entstehen und sich im Verlauf des Erwerbs entwickeln. Die Erweiterung der Perspektive betrifft auch die Anwendung auf Fragen zum Zusammenhang zwischen körperlichen, kognitiven, sozialen und entwicklungspsychologischen Prozessen und Zuständen, die durch einen übergreifenden Ansatz, etwa unter dem Schirm der Embodied Cognition, integrativ beantwortet werden könnten (vgl. Barsalou, 2008).

### **9.3. Konnektionismus und Embodied Language Processing**

Im Folgenden beschäftigen wir uns mit (eliminativ) konnektionistischen Modellen der situational verankerten Sprachverarbeitung auf lexikalischer und syntaktischer Ebene und des Spracherwerbs mit Fokus auf dem Bedeutungserwerb. Bereits im ersten Unterkapitel wurden immer wieder Modelle des Embodied Language Processing bzw. des Language Grounding erwähnt, die konnektionistische Architekturen miteinbeziehen. In diesem Unterkapitel besprechen wir Modelle, die einen spezifisch konnektionistischen Blick auf diese Themen werfen und den Konnektionismus nicht nur als Mittel zum Zweck, sondern als Quelle für die Annahme zentraler kognitiver Eigenschaften hervorheben. Wir haben in Kapitel 4.1 festgestellt, dass der Gegenstand Sprachverstehen mit sehr unterschiedlichen Ausgangspunkten und Zielen angegangen wird. In den folgenden Diskussionen werden wir die Überschneidungspunkte zu den Themen Language Grounding und Embodiment suchen. Dabei werden unter anderem die Integration von sprachlichen und nicht-sprachlichen Reizen als Fusionspunkt der Bedeutung sowie die Abkehr von propositionalen oder anderen symbolorientierten Organisationen von Informationen hervorgehoben.

Wir besprechen zunächst Modelle lexikalischer Verarbeitung, wobei das frühe Symbol-Grounding-Modell von Plunkett et al. (1992) im Vordergrund stehen wird. In diesem wird das Symbol-Grounding-Problem als Nebenprodukt des allgemeineren Symbol-Emergence-Problems diskutiert und eine konnektionistische Lösung vorgeschlagen. Aus ganz allgemeiner Perspektive bespricht Cangelosi (2005) das Thema und zeigt Zusammenhänge in der Herangehensweise verschiedener Ansätze in der Konzeption als Dual Route-Modelle. Auf SRNs und ihrer Interpretation als konnektionistische Implementierung eines perzeptuellen Symbolsystems (Joyce et al., 2003) basierend stellen Coventry et al. (2005) ein Modell für die Verarbeitung von Präpositionen vor. Des Weiteren besprechen wir kurz einige eher neurophysiologisch orientierte bzw. um neurobiologische Plausibilität bemühte Modelle (Li, Farkaš & MacWhinney, 2004; Wennekers et al., 2006; Garagnani et al., 2008). Im zweiten Abschnitt diskutieren wir Modelle syntaktischer Verarbeitung. Die Modelle wurden bereits in Kapitel 4.1 angesprochen und werden nun mit stärkerem Fokus auf Fragen der Embodied Cognition und der Sprachverankerung betrachtet. Speziell in den Ansätzen von (Altmann & Mirkovic, 2009), (Mayberry et al., 2009) und S. L. Frank et al. (2009) wird ein expliziter Bezug zu diesen Aspekten hergestellt. So schlagen zum Beispiel Frank und Kollegen ein Modell des Satzverstehens im Sinne der Aktivierung eines (analogen)

Situationsmodells vor (vgl. auch S. L. Frank et al., 2003b). Einen ähnlichen Anspruch mit Fokus auf Antizipationsprozessen verfolgen (Altmann & Mirkovic, 2009) in einem rein theoretischen Modell mit konnektionistischer Perspektive. In den gängigen Modellen des Sprachverstehens gibt es andererseits viele Gesichtspunkte zu berücksichtigen, die sich nicht sinnvoll mit den Ideen der Embodied Cognition verbinden lassen. Aus dieser beidseitigen Beleuchtung der Ansätze heraus formulieren wir in einem kritischen Überblick Kriterien, die konnektionistische Modelle des Embodied Language Processing erfüllen sollten.

#### 9.3.1. Modelle lexikalischer Verarbeitung

Ein gemeinsames Prinzip vieler konnektionistischer Ansätze zum Language Grounding ist die Idee der Dual-Route-Architektur (Cangelosi, 2005). Verschiedene Modalitäten werden über verschiedene Routen – verknüpfte Layer, die vom Input bis zum Output eine bestimmte Transformation leisten – realisiert, die an irgendeinem Punkt im Netzwerk in einen gemeinsamen Layer zusammengeführt werden (vgl. Abb. 9.1). Die Repräsentationsstruktur und auch die Transformationsaufgabe können also über die Routen sehr unterschiedlich sein. Im gemeinsamen Layer muss sich jedoch ein Repräsentationsformat finden, das trotz der Zusammenführung die Erfüllung aller verschiedenen Aufgaben der Routen über den geteilten Repräsentationsraum gewährleistet. Dieser Flaschenhals der Aktivierungsweiterleitung und Repräsentation dient als Grundlage der Informationsintegration verschiedener Modalitäten bzw. Informationssysteme. Über die gemeinsamen Repräsentationen bzw. durch die Möglichkeit der gegenseitigen Beeinflussung im gemeinsamen Layer dienen sich die verschiedenen Routen als gegenseitige Constraints. Zudem verweisen sie durch die Aktivierung der für beide relevanten Anteile der gemeinsamen Repräsentation aufeinander. Diese Prinzipien sind fundamental für das von uns entwickelte Modell, das wir in Kapitel 10 vorstellen werden.

Durch diese Eigenschaften wird die Dual-Route-Architektur zur Grundlage für Ansätze des Symbol oder Language Grounding. So kann etwa eine Route die visuelle und eine zweite Route die sprachliche Verarbeitung realisieren, die im gemeinsamen Layer in den visuellen Informationen und ihrer Verarbeitung gegroundet wird. Dabei können verschiedene Verarbeitungsmodalitäten simuliert werden:

- Situiertere Sprachverarbeitung bei der Darbietung von Beispielen über beide Informationskanäle, deren Verarbeitungsqualität innerhalb der Modalität und deren Einfluss auf die jeweils andere Modalität analysiert werden kann.
- Sprachverstehen bei der ausschließlichen Darbietung von Input über den Sprachkanal, wofür auf dem visuellen Output-Layer eine Reproduktion korrespondierender visueller Informationen geleistet wird, die als Aktivierung des Bedeutungsinhalts im Sinne der Simulation visueller Erfahrungen verstanden werden kann.
- Sprachproduktion bei der ausschließlichen Darbietung von Input über den visuellen Kanal, wofür auf dem sprachlichen Output-Layer eine Reproduktion der zur visuellen Information korrespondierenden Benennung geleistet wird.

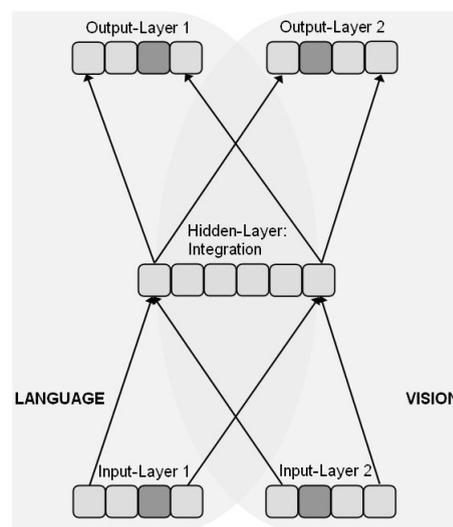


Abbildung 9.1.: Dual-Route-Architektur: Feedforward-Netzwerk (FFN) mit geteilten Input- und Output-Layern, auf denen unterschiedliche Repräsentationen dargeboten werden. Der Hidden-Layer dient als integrative Schnittstelle, in der gemeinsame interne Repräsentationen gebildet werden.

**Ein früher Ansatz** Die Abbildung von Form auf Bedeutung über verschiedene Input-Output-Repräsentationen, die sich gegenseitig beeinflussen können, wurde bereits im „Cats, Dogs and Bagels“-Modell (McClelland & Rumelhart, 1985) thematisiert. Das Ziel dieses Modells ist es zu zeigen, wie mit einem konnektionistischen System ein verteilt repräsentierendes Gedächtnissystem modelliert werden kann, das ohne die Annahme verschiedener Module die Repräsentation von allgemeinen und spezifischen Informationen ermöglicht. Das Modell ist als einschichtiger Auto-Assoziator realisiert. Man kann hier also noch nicht von einem Modell im Sinne der Dual-Route-Architektur sprechen. Die Autoren beschreiben, wie sich in diesem Modell funktionale Äquivalente abstrakter Repräsentationen – Prototypen, Logogene und regelhaftes Verhalten – aus Überlagerungen von erfahrungsba-sierten Gedächtnisspuren entwickeln. Dabei kann es ohne Modularisierung oder explizite Type-Token-Distinktion und trotz der Generalisierung über die Input-Ereignisse sowohl allgemeine (z.B. die Kategorie Hund) als auch spezifische Informationen (z.B. der konkrete Nachbarshund Waldi) erhalten und reproduzieren.

**Symbolgenese in Dual-Route-Netzwerken** Ein sehr ähnliches Modell stellen Plunkett et al. (1992) vor: In diesem werden die Grundprinzipien des Modells von (McClelland & Rumelhart, 1985) übernommen und auf die Frage des Symbol Grounding und des Vokabelerwerbs übertragen. Plunkett und Kollegen problematisieren klassisch-funktionalistische Ansätze, und stellen diesen eine Alternative auf Grundlage künstlicher neuronaler Netzwerke gegenüber. Anhand dieser beschreiben sie, wie das Symbol-Grounding-Problem in einem epigenetisch-entwicklungspsychologischen Ansatz als Nebenprodukt der Lösung des Symbol-Emergence-Problems gelöst wird. Als großer Vorteil der konnektionistischen Ansätze wird ihre Unabhängigkeit von der Voraussetzung der funktionalen Eineindeutigkeit thematisiert – einer Voraussetzung für Abbildungen, die in klassischen Modellen erfüllt

werden muss. Diese kann für die Aufgabe der Referenzbildung aber nicht gewährleistet werden, denn Referenz zeichnet sich in der echten Welt dadurch aus, dass ihre Abbildungen mehrdeutig, nicht-atomar und kontextabhängig sind.

Für die Modellierung schlagen Plunkett et al. (1992) eine Dual-Route-Architektur vor: Das Modell ist als mehrschichtiges Feedforward-Netzwerk realisiert, wobei die Input- und Output-Layer aus jeweils zwei getrennten Teil-Layern bestehen, die verschiedene Modalitäten repräsentieren: Eine Verarbeitungsrouten realisiert die visuelle Verarbeitung, die zweite die linguistische Verarbeitung. Im Training werden wie im „Cats, Dogs and Bagels“-Modell korrespondierende visuelle und linguistische Informationen zusammen präsentiert. Beide Routen führen über einen gemeinsamen Kompressions-Layer, in dem eine gemeinsame integrierte Repräsentation entwickelt werden muss. Da im Input und im Output jeweils die gleichen Informationen präsentiert werden, entspricht das Modell einem auto-assoziativen Ansatz.<sup>12</sup> Die linguistische Repräsentation entspricht orthogonalen lokalistischen Vektoren. Die visuelle Repräsentation entspricht Verzerrungen zufälliger Punktmuster auf einer zweidimensionalen Retina, wobei das Trainingsset keines der prototypischen Punktmuster enthält, sondern nur die Verzerrungen, die den erfahrbaren Instanzen des Prototyps entsprechen. Nachbarschaftsinformationen werden in der zweidimensionalen Retina durch rezeptive Felder auf dem Input-Layer realisiert.

Mit dem Modell wird die Symbolgenese als situational verankerte Kategorienbildung simuliert. Der Vokabelerwerb erfolgt hier durch die Dekontextualisierung sprachlich benannter Konzepte von ihrem kontextuellen Setting. Die Entwicklung von Symbolen erfolgt dabei nicht als Entkopplung symbolischer Repräsentationen von ihrem subsymbolischen Substrat, sondern als Konstruktion erweiterter funktionaler Organisation, die nach wie vor subsymbolischer Natur ist. Damit ist die situationale Verankerung von Sprache in der Wahrnehmung bereits inbegriffen, die Lösung des Symbol-Grounding-Problems erfolgt als Nebenprodukt (vgl. Plunkett et al., 1992). Auch aus empirisch-entwicklungspsychologischer Sicht sind die Simulationsergebnisse interessant. In der Performanz der Netzwerke zeigen sich vier zentrale Eigenschaften des Vokabelerwerbs, die aus empirischen Studien bekannt sind: (a) Prototypen-Effekte bei der Kategorisierung, (b) Asymmetrien zwischen Verstehen und Produktion,<sup>13</sup> (c) der Vokabelspurt als sprunghafter Anstieg gelernter Vokabeln,<sup>14</sup> und (d) Effekte der Über- und Untergeneralisierung für Benennungen.

Eine berechtigte Kritik an diesem Ansatz äußert Davis (2002): Im Modell wird der Vokabelerwerb als kontrollierte Definition mit klarer Zuweisung der abzubildenden Bedeutung dargestellt. Dies lässt viele Aspekte und Probleme der tatsächlich gegebenen Situation beim Spracherwerb unberücksichtigt. Vor allem das Problem der Etablierung von Eins-zu-eins-Korrespondenzen aus allen möglichen Viele-zu-viele-Abbildungen zwischen

---

<sup>12</sup>Der Trainingsaufbau ist allerdings etwas komplizierter gestaltet: Für jedes Beispielpaar erfolgen immer drei Aktivierungsdurchläufe: Jeweils ein einzelner Durchlauf für die visuelle und die linguistische Route und schließlich ein gemeinsamer Durchlauf über beide Routen.

<sup>13</sup>Allerdings schränken Plunkett et al. (1992) berechtigterweise ein, dass diese Asymmetrie auch auf die zugrundegelegten Repräsentationsformate für die Darbietung der Beispiele zurückgeführt werden könnten. Die höhere Komplexität von visuellen gegenüber sprachlichen Stimuli ist eher eine explizite Annahme der Modellierer, die in der gewählten Form allerdings nicht gänzlich unbegründet erfolgt.

<sup>14</sup>Für den Vokabelspurt wurde zudem vorhergesagt, dass er nicht nur für Sprachproduktion, sondern auch für das Sprachverstehen stattfinden müsste.

mehreren Wörtern einer Äußerung und verschiedenen Situationsaspekten, das Lösen wie zum Beispiel situationsübergreifendes Lernen erfordert, bleibt hier unberücksichtigt.

**Erwerb abstrakter Konzepte** Konnektionistische Dual-Route-Architekturen stellen auch die Basis für Modellierungen zur Symbolic-Theory-Hypothese und dem hier angenommenen Mechanismus des Grounding Transfers (Cangelosi & Harnad, 2000; Cangelosi et al., 2002). Konnektionistische Modelle sind geeignete Kandidaten für die hierfür erforderliche kategoriale Perzeption. Cangelosi, Greco und Harnad (2000) beschreiben ein Modell, das Objekte mit verschiedener Form und Farbe auf einer zweidimensionalen Retina kategorisiert. Auf der sprachlichen Route werden korrespondierende lexikalische Beschreibungen dieser Oberflächenmerkmale dargeboten. In einem späteren Lernschritt werden neue Kategorien ohne direkten Bezug zu visuellem Input aus der bloßen Beschreibung durch neue Kombinationen lexikalischer Beschreibungen von Oberflächenmerkmalen erworben. Diese Kategorien können im Test mit rein visuellem Input entsprechender Objekte geprüft werden. Die Objekte werden erkannt, obwohl sie nie gesehen wurden, da die neuen Kategorien auf Grundlage sprachlicher Symbole erworben werden, die im visuellen System gegroundet sind. Das Modell simuliert das Erlernen von Konzepten höherer Abstraktion auf Basis gegroundeter Konzepte.

**Benennung komplexer visueller Ereignisse** Ein weiteres interessantes Modell ist das Functional Geometric Framework von Coventry et al. (2005): In diesem werden Präpositionen auf Grundlage räumlicher und funktionaler Gesichtspunkte auf perzeptuelle Repräsentationen komplexer visueller Ereignisse abgebildet (vgl. auch Cangelosi et al., 2004). In den Ansätzen von Regier (1995) und Regier und Carlson (2001) werden für die sprachliche Benennung räumlicher Semantik nur räumlich-geometrische Aspekte berücksichtigt. Wie Coventry, Prat-Sala und Richards (2001) aber in einer Rating-Studie zu den Präpositionen *over*, *under*, *above* und *below* feststellen, beziehen sich die ersten beiden Präpositionen stärker auf funktionale Aspekte als auf räumliche Aspekte der Relation zwischen Objekten. Ein Regenschirm befindet sich also nur dann auf der Position *over*, wenn er seine Funktion erfüllt, den Träger vor Regen zu schützen – auch wenn er sich bei schräg fallendem Regen in der Tat räumlich vor seinem Träger befindet. So beschreibt *over* also eher einen funktionalen und *above* eher einen räumlichen Aspekt. Darüber hinaus existieren aber noch weitere Faktoren, die räumliche Semantik bestimmen, wie etwa objektspezifisches Wissen sowie Wissen über die Interaktionsdynamik von Objekten. Solche Unterschiede der relevanten Faktoren sollen im Functional Geometric Framework berücksichtigt werden.

Grundlage der Verarbeitung von Ereignissen ist das Connectionist Perceptual Symbol System Network (CPSSN, Joyce et al., 2003), mit dem dynamische Ereignissequenzen erfasst, komprimiert und konzeptualisiert werden können. Das CPSSN ist als SRN mit Vorhersageaufgabe für die sukzessiven Teilereignisse der Ereignissequenz realisiert. Wir werden noch einmal in Kapitel 9.2 auf diese Architektur und die daraus gezogenen Schlüsse zu sprechen kommen. Im Functional Geometric Framework wird die visuelle Verarbeitung durch eine hybride Architektur geleistet, in der visuelle Routinen (Ullman, 1996) mit

SRNs (im Sinne des CPSSN, Joyce et al., 2003) kombiniert werden.<sup>15</sup> Die Integration visueller und linguistischer Informationen wird durch eine Dual-Route-Architektur geleistet: In dieser werden auf der visuellen Route statische Repräsentationen der Ereignisse abgebildet, die aus den Hidden-Layern der SRNs für den jeweils letzten Zeitschritt eines Ereignisses abgeleitet wurden. Auf der sprachlichen linguistischen Route wird die jeweils korrespondierende Beschreibungen als orthogonale Repräsentation der Objektbenennung und einer passenden Präposition dargeboten. Die zur visuellen Szene korrespondierenden lexikalischen Beschreibungen entsprechen in ihrer Häufigkeit den Verhältnissen, die in der Rating-Studie gezeigt wurden.<sup>16</sup> In Sprachproduktionstests erfasst das Modell neue Ereignisse und generiert eine Verteilung der möglichen sprachlichen Beschreibungen. Die Aktivierungsverteilung für die Präpositionen weisen dabei Verhältnisse auf, die mit den Befunden empirischer Rating-Studien übereinstimmen. Zudem trifft das Modell Vorhersagen zur Beurteilung nicht ausgeführter Ereignisse: Dazu werden dem Modell nur unvollständige initiale Ereignissequenzen gegeben, aus denen das Modell über eine Feedback-Schleife Handlungsabläufe generieren soll (vgl. zu einem ähnlichen Ansatz Botvinick & Plaut, 2004). Aus diesen können wiederum visuelle Repräsentationen des Endzustands und korrespondierende sprachliche Beschreibungen abgeleitet werden. Die Ergebnisse wurden mit einem Folgeexperiment überprüft: Dabei zeigte sich eine hohe Übereinstimmung zwischen Simulationsdaten und Rating-Verhältnissen (Coventry et al., 2005).

Problematisch ist an diesem Modell, dass eine hohe Anzahl an Zwischenschritten zur Integration der Module erforderlich ist, die viel Handarbeit durch die Modellierer erfordert, und dass die Module getrennt trainiert werden. In diesem Zusammenhang erscheint die linguistische Route als bloßes Anhängsel an eine statische Schnittstelle.

#### 9.3.1.1. Neuronal plausible Modelle

In den letzten Jahren wurden vermehrt Modelle vorgeschlagen, deren Hauptaugenmerk darauf liegt, neuronal plausibel zu sein. Auch in diesen Modellen geht es um die Konzeptualisierung und Integration verschiedener Informationen. Allerdings wird dabei auf klassische Konzepte aus der Tradition des Parallel Distributed Processing verzichtet, deren Plausibilität aus neurobiologischer Perspektive bemängelt wird. Möglicherweise kann durch die bessere neuropsychologische Vergleichbarkeit von Modell und Empirie eine weitere Brücke zwischen Konnektionismus und Embodied Cognition geschlagen werden.

Nyamapfene und Ahmad (2007) schlagen ein Modell vor, das den Ansatz von Plunkett et al. (1992) aufgreift. In diesem wird die Verwendung des Backpropagation-Algorithmus durch selbstorganisierende Lernverfahren ohne Lehrersignal ersetzt. Nyamapfene und Kollegen zeigen, dass selbstorganisierte Netzwerke ähnliche Ergebnisse hervorbringen wie das Vorgängermodell. Allerdings bestätigen sie die Einschränkungen der Aussagekraft solcher

---

<sup>15</sup>Da in den dynamischen Szenen mehrere Objekte und Aspekte abgebildet werden, mit dem visuellen Modul aber die autonome Segmentierung und Identifizierung verschiedener Objekte nicht möglich ist, benötigt das Modell von Cangelosi et al. (2004) so viele SRNs wie Objekte und Aspekte. Die hier verwendeten Szenen benötigen drei SRNs für ein Quell-Objekt, ein Ziel-Objekt und ein bewegliches Objekt.

<sup>16</sup>Allerdings werden immer eindeutige Beschreibungen dargeboten. Die relative Häufigkeit alternativer funktionaler oder geometrischer Beschreibungen entsprechend der Verhältnisse aus den Rating-Studien wird über mehrere Beispiele hinweg in der Präsentationsfrequenz erfasst.

Modelle des Spracherwerbs in der Ein-Wort-Phase, die auch von Davis (2002) bemängelt wurde. Zudem kritisieren sie, dass in diesen Modellen in der Regel ignoriert wird, dass Kinder keine passiven Rezeptoren von Input aus der Umwelt sind, sondern selbst aktiv mit der Umwelt interagieren und kommunikative Intentionen besitzen. Einen stärkeren neuro- und entwicklungspsychologischen Ansatz wählen Li et al. (2004) mit einem Modell der frühen lexikalischen Entwicklung. In ihrem DevLex-Modell werden zwei wachsende selbstorganisierende Karten kombiniert (vgl. auch Farkaš & Li, 2002). Gegenstand der Simulationen sind drei Aspekte des frühen Erwerbs: Die Aufteilung verschiedener Wortklassen in der kortikalen Topographie, dynamische Effekte des Vokabelerwerbs (der sogenannte Vokabelspurt und damit assoziierte temporäre Fehlleistungen bei der Benennung) und Effekte des Erwerbsalters (welche Wörter werden wann in welcher Reihenfolge erworben). Während einerseits die übliche Kritik der vereinfachten Abbildungssituation gilt, kann das Modell andererseits eine realistische Skalierung des zu verarbeitenden Spektrums sprachlicher Informationen leisten.

Die im vorhergehenden Absatz vorgestellten Modelle beschränken den Anspruch neuro-naler Plausibilität auf die Verwendung entsprechender unüberwachter Lernverfahren (vgl. für ähnliche Ansätze Hadley et al., 2001; Candliss & Fiez, 2002; Gorrell & Webb, 2006). Einen Schritt weiter geht der Modellierungsansatz, der von Wennekers et al. (2006) und Garagnani et al. (2008) entwickelt wurde: In diesem stehen neben der Verwendung realistischer Lernverfahren die Architektur, der neurofunktionale Bezug und die Ableitung von Vorhersagen für empirische Befunde aus bildgebenden neurowissenschaftlichen Verfahren im Vordergrund. Mit den Modellen wird die Neurophysiologie spezifischer Systeme im Gehirn nachgebildet, etwa der für die Sprachverarbeitung zuständige Kortex. Wennekers et al. (2006) beschreiben die Entwicklung von hebbischen Cell-Assemblies, stark verknüpften funktionalen Gruppierungen von Aktivierungen von Neuronen unterschiedlicher kortikaler Areale (vgl. Pulvermüller, 1999). Im Modell werden die kortikal übergreifenden Cell-Assemblies auf Grundlage hebbischer Lernverfahren durch die drei Mechanismen Hetero-Assoziation, Auto-Assoziation und konditionale Assoziation gebildet. Das Modell simuliert Objektbenennung als Objektkategorisierung in visuellen Arealen und daraus generierten motorischen Sequenzen für die phonetische Realisierung der Benennung beim frühen Vokabelerwerb. Dabei lässt sich die Formierung von Cell-Assemblies durch synchrone, über verschiedene kortikale Areale verteilte Aktivierungen beobachten. Garagnani et al. (2008) demonstrieren die Simulation der Verarbeitungsunterschiede für bedeutungsvolle Wörter gegenüber bedeutungslosen Pseudowörtern (vgl. auch Garagnani, Shtyrov & Pulvermüller, 2004). Die Performanz des Modells trifft mehrere kritische Vorhersagen, für die sich in nachfolgenden empirischen neurofunktionalen Studien positive Befunde ergaben. Insbesondere werden Effekte der Aufmerksamkeit<sup>17</sup> auf neurofunktionale Effekte wie Latenz und Amplitude vorhergesagt und begründet: Durch ihre starke interne Konnektivität entwickeln sensomotorische Muster für bedeutungsvolle Wörter eine Aktivierungsdynamik, die von Aufmerksamkeitseinschränkungen weitgehend unbeeinflusst bleibt. Die Verarbeitung von Pseudowörtern hingegen ist direkt von der Verfügbarkeit der Aufmerksamkeitsressourcen bedingt, da sie verschiedene konkurrierende Repräsentationen

---

<sup>17</sup>Differenzen der Aufmerksamkeit werden als unterschiedliche Levels verfügbarer Aufmerksamkeitsressourcen modelliert, indem das globale Level der Inhibition verändert wird, das die Gesamtmenge der im Netzwerk möglichen Aktivierung bestimmt.

aktiviert, die durch inhibitorische Prozesse kontrolliert werden. Befunde entsprechender neurofunktionaler Korrelate decken sich mit dieser Erklärung.

#### 9.3.1.2. Repräsentationale Spezifität

Welche Position wird mit den konnektionistischen Ansätzen zur Frage der Spezifität semantischer Repräsentationen und darauf bezogener Verarbeitungsmechanismen für verschiedene Modalitäten eingenommen? Welche Natur wird für die Repräsentationen und Mechanismen angenommen: Erfolgt die semantische Verarbeitung modalitätsspezifisch auf unterschiedlichen Repräsentationen und mit spezifischen Constraints? Oder ist hierfür eine übergreifende Klasse von Mechanismen und Constraints zuständig, die auf einer gemeinsamen Repräsentationsebene operiert und zwischen verschiedenen Modalitäten vermittelt bzw. über diese abstrahiert? Wir haben diese Fragen bereits für verschiedene hybride Modelle (Partridge, 1995; Schank & Fano, 1995; Robert-Ribes et al., 1995) angesprochen. In konnektionistischen Ansätzen werden hierauf zumeist keine expliziten Antworten gegeben. In gleichem Maße bleiben aber auch in Ansätzen der Embodied Cognition die Antworten eher vage.

Plaut (2002) stellt eine Studie vor, in der folgende Frage explizit thematisiert wird: Ist semantisches Wissen in einem einzelnen amodalen (im Sinne von: über Modalitäten abstrahierenden) System oder in multiplen modalitätsspezifischen Subsystemen organisiert? In diesem Zusammenhang diskutiert Plaut Argumente von Vertretern unitärer und multipler Semantiktheorien. Gegenstand der Studie ist eine konnektionistische Modellierung optischer Aphasie. Dieser Gegenstand wurde deshalb gewählt, weil er Theorien eines singulären semantischen Systems Probleme bereitet: Für modalitätsspezifische Aphasien ist bekannt, dass Patienten selektive Ausfälle bei der Benennung von assoziierten Sachverhalten haben. Wenn etwa eine optische Aphasie durch eine lokale Schädigung des visuellen Systems an der medialen Seite des Occipitallappens vorliegt, werden visuell dargebotene Objekte zwar erkannt, können aber nicht benannt werden. Umgekehrt gibt es empirische Einwände gegen Theorien, die ein multiples modalitätsspezifisches semantisches System annehmen: Trotz der Einschränkungen bei der Benennung von Objekten können visuell wahrgenommene Handlungen relativ problemlos benannt werden. Gäbe es ein disjunktes verbales und visuelles semantisches System, und wären diese Systeme durch eine Hirnschädigung getrennt, so sollte eine solche Dissoziation nicht möglich sein. Mit dem konnektionistischen Ansatz stellt Plaut den entgegengesetzten unitären und multiplen Ansätzen eine Theorie graduell modalitätsspezifischer Spezialisierung in der semantischen Verarbeitung gegenüber, die einen Mittelweg als Lösung der empirischen Datenlage sucht (vgl. auch Rogers & Plaut, 2002). Nach dieser Theorie entwickeln sich semantische Repräsentationen unter dem Druck, die Vermittlung zwischen verschiedenen Modalitäten und Aufgaben leisten zu müssen (Plaut, 2002, S. 607): „[...] semantics is a learned, internal representation that develops under the pressure of performing a variety of tasks involving various input and output modalities [...]“. Das semantische System muss eine graduelle funktionale Spezialisierung entwickeln, um die Vermittlung zwischen verschiedenen Modalitäten leisten zu können. Damit verändert sich die Fragestellung hinsichtlich der Natur semantischer Repräsentationen: Welcher Natur sind die funktionalen Spezialisierungen und wie entwickeln sie sich? Plaut identifiziert zwei grundlegende Faktoren, die die Eigen-

schaften der Spezialisierung prägen: (a) Unterschiede der relativen Systematizität zwischen verschiedenen Aufgaben bzw. erforderlichen Transformationen, und (b) ein Einfluss der neuronalen Topographie auf das Lernen und die Prägung des Repräsentationsraums durch die Bevorzugung kurzer Verbindungsstrecken.

Für die Modellierung wurde ein kontinuierlich rekurrentes Attraktornetzwerk verwendet, in dem visueller oder taktiler Input auf motorischen (Handlung) und/oder phonologischen (Sprache) Output abgebildet wird. Das semantische System ist als zweidimensional aufgebauter Hidden-Layer konzipiert, der mit den genannten Modalitäten verknüpft ist. Hier finden die kritischen Prozesse der Vermittlung zwischen verschiedenen Modalitäten und Aufgaben statt. Hinzu kommen zwei Aufgaben-Units, durch die bestimmt wird, ob das Netzwerk Objektbenennung (phonologischer Output für das wahrgenommene Objekt) oder Handlung und Handlungsbenennung (phonologischer Output für die zum Objekt assoziierte Handlung und motorischer Output für die entsprechende Handlung) leisten soll. Das Netzwerk zeigt starke Bezüge zum Ansatz der Dual-Route-Architektur, die hier um eine Gating-Komponente (die beiden Aufgaben-Units) und um topographische Eigenschaften erweitert wird. Die Repräsentationen des phonologischen Inputs wurden über phonetische Features und Konsonant-Vokal-Slots realisiert, für die anderen Modalitäten wurden Zufalls-Punktmuster für fünf Prototypen mit jeweils zwanzig Exemplaren (verzerrte Ableitungen) erstellt. Der Einfluss der Modalitäten auf verschiedene topographische Positionen der Repräsentationen im Hidden-Layer wird durch spezielle Lernverfahren von ihrer Distanz zu diesen Positionen abhängig gemacht. Je näher die jeweilige Modalität zur Position ist, desto größer ist ihr Einfluss auf die jeweilige Repräsentation und desto modalitätsspezifischer ist dieser Anteil an der Repräsentation. Positionen mit modalitätsübergreifend vergleichbarer Distanz hingegen operieren mit weitgehend amodaler Funktionalität. Durch diese Eigenschaft erhalten die Repräsentationen im System eine graduelle Spezifität und zeigen zugleich Effekte semantischer Ähnlichkeit. Abhängig von der Lokalisation von Läsionen sowohl innerhalb der semantischen Repräsentationen als auch über die modalen Verknüpfungen können verschiedene Aphasien simuliert werden – etwa optische Aphasie, visuelle Agnosie oder anomische Aphasie. Dabei werden innerhalb eines Systems ohne weitere spezielle Anpassungen spezifische Verhaltensmuster und Dissoziationen unterscheidbarer Aphasien reproduziert.<sup>18</sup>

Während in diesem einen vorgestellten Modell eine spezifische Antwort gegeben wurde, muss im Überblick über verschiedene konnektionistische Modelle kritisch hinterfragt werden, ob sich die Frage der Natur des semantischen Systems für diese Modellklasse sinnvoll beantworten lässt. Einerseits unterscheiden sich konnektionistische Ansätze ja gerade durch die Analogizität ihrer Repräsentationen zur Struktur der dargebotenen Beispiele. Andererseits ist anzunehmen, dass die Strukturen verschiedener Modalitäten in irgendeiner Form ein gemeinsames Format finden müssen, um sinnvoll assoziiert und integriert werden zu können. Und hinsichtlich der Mechanismen: Einerseits wirken in einfachen konnektionistischen Netzwerken übergreifend die gleichen Aktivierungs-, Hemmungs- und

---

<sup>18</sup>Ähnliche konnektionistische Modelle werden von Devlin, Gonnerman, Andersen und Seidenberg (1998) sowie von Farah und McClelland (1991) für kategorienspezifische semantische Störungen vorgeschlagen. Im Modell von Devlin und Kollegen etwa werden Dissoziationen verschiedener Kategorien mit einer entsprechenden Argumentation innerhalb eines semantischen Systems anhand von Unterschieden zwischen Objekten hinsichtlich perzeptueller und funktionaler Eigenschaften und anhand einer topographischen Organisation dieser Eigenschaften erklärt.

Lernmechanismen, andererseits erfolgt die Entwicklung von Constraints modalitäts- und aufgabenspezifisch, um die erforderlichen Transformationen zu leisten. Es fragt sich: Entwickeln sich an den Schnittstellen der Integration verschiedener Modalitäten übergreifende Constraints, die über die Modalitäten hinweg abstrahieren? Die Frage nach der Natur der semantischen Repräsentationen in konnektionistischen Netzwerken hängt aber letztendlich auch stark davon ab, welchen Aspekt der Netzwerke man als Substrat der semantischen Repräsentation verstehen möchte.<sup>19</sup> In Ansätzen mit Dual-Route-Architektur werden zumeist die Aktivierungsmuster des integrierenden Hidden-Layers, über den die Abhängigkeiten und gegenseitigen Constraints zwischen verschiedenen Informationssystemen ausgehandelt werden, als Kern des semantischen Systems und als Substrat der Bedeutung verstanden (vgl. z.B. Plunkett et al., 1992). Eine ähnliche Position vertreten Plaut (2002) und Rogers und Plaut (2002). Aus dieser Perspektive vertreten konnektionistische Modelle des lexikalischen Bedeutungserwerbs am ehesten die Position der graduell modalitätsspezifischen Spezialisierung. Damit bieten sie auch einen Ansatz für die Theoriebildung der Embodied Cognition.

### 9.3.2. Modelle syntaktischer Verarbeitung

Im folgenden Unterkapitel besprechen wir Modelle der syntaktischen Verarbeitung aus der Perspektive des Embodied Language Processing. Dem ist vorwegzunehmen, dass es eigentlich keine konnektionistischen Ansätze gibt, die syntaktische Verarbeitung explizit im Sinne des Embodied Language Processing modellieren. Vielmehr gibt es verschiedene Modelle, die hinsichtlich ihrer Eigenschaften auf die Annahmen der Embodied Cognition bezogen wurden. Wir werfen nun einen genaueren Blick darauf, inwiefern mit diesen Modellen welche Konzepte der Embodied Cognition implementiert werden. Wir besprechen die Umsetzung der Dual-Route-Architektur in Kombination mit SRNs, das CIA-Net (Mayberry et al., 2005), das Modell von Dominey (2003), und mit besonderem Gewicht die verschiedenen Versionen des DSS-Modells (S. L. Frank et al., 2003a, 2009). Umgekehrt müssen wir fragen, inwiefern andere konnektionistische Modelle des Satzverstehens den Annahmen gerecht werden, bzw. inwiefern sie davon abweichen. Können etwa das SG-Modell (McClelland et al., 1989; St. John, 1992), das CSCP-Modell (Rohde, 2002) das INSOMNet (Mayberry, 2003) oder das Modell von (Desai, 2001) als Modellierungsansatz für Theorien des Embodied Sentence Processing herangezogen werden?

#### 9.3.2.1. Annäherungen an die Embodied Cognition

Bevor wir die Eignung der verschiedenen konnektionistischen Modelle als Modellierungsansätze für das Satzverstehen im Embodiment-Ansatz diskutieren, müssen wir zunächst wie bereits in Kapitel 4 feststellen, dass die meisten Ansätze gänzlich andere Fragen behandeln – etwa die Simulation psycholinguistischer Daten, die Entwicklung eines ernstzunehmenden konnektionistischen Parsers, die Untersuchung semantischer Systematizität

---

<sup>19</sup>Dieses kann gleichermaßen in den Aktivierungen der assoziativen modalitätsübergreifenden Schnittstellen, den Gewichten von und zu diesen oder aber verteilt über zusammenhängende Aktivierungsmuster im gesamten Netzwerk gesucht werden. Je nachdem, welchen dieser Aspekte man als semantisches System betrachten möchte, wird die Antwort unterschiedlich ausfallen (vgl. Kap. 4.5).

in konnektionistischen Netzwerken oder die Skalierbarkeit dieser Modelle auf die Erfassung eines Korpus realistischer Größe und Komplexität. Und in erster Linie müssen die Modelle natürlich in diesen Domänen evaluiert werden, für die sie entwickelt wurden. Dennoch wurden in verschiedenen Modellen klare Bezüge zur Embodied Cognition hergestellt. Dies betrifft vor allem Aspekte der Informationsintegration verschiedener Domänen oder Modalitäten, des situierten Erwerbs und der situierten Verarbeitung von Sprache. Des Weiteren betrifft es die Frage der Enkodierung der semantischen Informationen und die dahinterstehenden Annahmen zur Entwicklung und Natur semantischer Repräsentationen.

**Integration verschiedener Domänen** Eine sehr geradlinige Möglichkeit, die wir bereits mit den lexikalischen Modellen vorgestellt haben, bietet der Ansatz der Informationsintegration und des Language Grounding über eine Dual-Route-Architektur. Eine Erweiterung zur Verarbeitung sequenzieller Strukturen etwa mit SRNs ermöglicht die Simulation von Satzverarbeitung. Die Integration zweier Domänen über einen rekurrenten Hidden-Layer wird zum Beispiel von Dienes et al. (1999) vorgeschlagen und von Altmann und Mirkovic (2009) auf Fragen der Integration verschiedener Modalitäten in einem prädiktiven System übertragen. Die Eigenschaften eines solchen Systems leiten sie direkt aus den Eigenschaften von (erweiterten) SRNs ab. Dabei spielt die mögliche Überlappung der durch ein wahrgenommenes Ereignis aktivierten Repräsentationen eine große Rolle für die Antizipation des weiteren Verlaufs des Ereignisses sowie des sich entwickelnden Satzes. Dabei unterstellen die Autoren ein multimodales und dabei unitäres System, in dem Erwerb, Verarbeitung und Repräsentation von Wissen unterschiedlicher Domänen auf den gleichen Mechanismen beruhen. Dadurch sind in diesem System Äußerungsbedeutungen, szenisch-visuelle Informationen und linguistische Erwartungen nicht unterscheidbar. Diese Nähe zu Konzeptionen des Embodied Language Processing verdeutlichen Altmann und Mirkovic (2009, S. 38) folgendermaßen:

Within the framework outlined here, simulation can be equated with changes to the internal state of the substrate shared between language and different sensorimotor domains which enable predictions regarding the changes in real-world (or indeed, bodily) states that are entailed by the event structures described by that language.

Der Ansatz bleibt allerdings nur eine Rahmentheorie ohne tatsächliche Implementierung, und folglich wurden auch keine Details zur Umsetzung angedacht. Unabhängig davon wurde von Weldle, Konieczny et al. (2009a) ein ähnliches Modell vorgeschlagen, das bei Abweichung einiger Spezifikationen und Grundannahmen ebenfalls eine Verbindung der Dual-Route-Architektur mit der SRN-Architektur implementiert und konkreter auf darin implementierte Aspekte der Embodied Cognition eingeht. Dieses Modell werden wir im Folgekapitel 10 ausführlich vorstellen.

Ein ähnlicher Ansatz der Integration verschiedener Informationen über ein SRN wurde mit dem CIA-Net (Mayberry et al., 2005, 2009) vorgeschlagen. Anstatt wie in der Dual-Route-Architektur werden dabei die beiden Kanäle – visueller und sprachlicher Input – nur auf eine gemeinsame Repräsentation abgebildet, die einen reduzierten Ausschnitt der visuellen Situation erfasst. Damit wird zum einen visuelle Aufmerksamkeit modelliert, zum anderen entspricht der Erwerb der integrierten und interagierenden Informationen einer linguistischen Indexierung nicht-linguistischer erfahrungsbasierter modaler Repräsentatio-

nen, wie sie zum Beispiel von Glenberg und Robertson (2000) vorgeschlagen wurde. Die Autoren weisen darauf hin, dass das Modell in diesem Sinne als Simulation des situierten Spracherwerbs verstanden werden kann, und dabei eine Vereinigung kookkurrenzbasierter (etwa Landauer et al., 1998) und situierter (etwa Barsalou, 1999) Perspektiven auf Sprache eröffnet (vgl. Mayberry et al., 2009, S. 489).

**Analogizität der Repräsentationen** Ein weiterer Ausgangspunkt, der in Beziehung zur Embodied Cognition gebracht werden kann, ist die Implementierung einer direkten Abbildung von sprachlichen Strukturen auf Repräsentationen szenisch-situationalen Wissens im Sinne einer linguistisch induzierten mentalen Simulation einer Situation. Im Unterschied zu den vorangegangenen Ansätzen geht es dabei nicht um die Integration verschiedener Input-Modalitäten, sondern ausschließlich um die Abbildung sprachlicher Sequenzen auf eine statische Situationsrepräsentation. Allerdings ist dies ein Ausgangspunkt, der im Grunde in allen konnektionistischen Modellen des Satzverstehens in der ein oder anderen Form gewählt wird. Was einen entsprechenden Ansatz für eine Embodiment-Perspektive relevant und interessant macht, ist die Wahl der semantischen Repräsentation und deren Vergleichbarkeit mit Annahmen der Embodied Cognition zur Repräsentation konzeptuellen Wissens.

In genau diesem Zusammenhang steht das DSS-Modell (S. L. Frank et al., 2003b, 2008, 2009). Frank und Kollegen versuchen hier, eine explizit nicht-propositionale Repräsentation zu entwickeln. Dies halten sie unter anderem deshalb für erstrebenswert, weil Propositionen Eigenschaften und Beschränkungen bei der Bildung von Inferenzen während des Sprachverstehens nahelegen, die sich nicht mit neueren Befunden speziell in der Tradition des Embodied Language Processing decken. Zudem vermeidet die Verwendung einer nicht-propositionalen Repräsentation den Verdacht der Implementierung eines klassisch-symbolischen Systems. Sprachverstehen entspricht in ihrem Ansatz nicht der Konstruktion von propositionalen Strukturen, sondern der Simulation bzw. der Aktivierung eines (mentalen) Modells der beschriebenen Szene (vergleichbar etwa mit der Konzeption von Zwaan, 2004). Die dabei entstehende Repräsentation ist in der Erfahrung des Verstehenden mit der Welt begründet. Sie beinhaltet das damit verbundene multimodale Wissen und steht in Analogie zu den Strukturen der erfahrenen Welt. Sie ist analog und multimodal (vgl. Barsalou, 1999) und ermöglicht deshalb beim Sprachverstehen, direkt Inferenzen zu bilden. Für das DSS-Modell wird konstatiert, dass es in diesem Sinne zwar nicht die Modalität, doch aber die Analogizität der Repräsentationen erfasst (S. L. Frank et al., 2009, S. 360). Die Analogizität zu den Strukturen der repräsentierten Welt wird auch als Schlüssel zur Systematizität der Verarbeitung verstanden (S. 376): „The origin of systematicity should be sought in the cognitive system’s embeddedness in the world rather than in inherent properties of the system itself.“

#### 9.3.2.2. Eine kritische Eignungsprüfung

Wo liegen die Grenzen der Eignung verschiedener konnektionistischer Modelle des Sprachverstehens für die Modellierung von Aspekten der Embodied Cognition? Die mitunter andere Zielsetzung der Modelle ändert nichts an der Beurteilung, ob sie sich für die Simulation des Sprachverstehens im Embodiment-Ansatz eignen oder nicht. Klare Diffe-

renzierungen wurden in dieser Hinsicht bislang kaum geäußert. Dabei gibt es deutliche Einschränkungen, die die gewählten Repräsentationsformen, die Art des Erwerbs und die Art der Integration von Informationen betreffen.

**Propositionale Repräsentationen** Die Realisierung der Repräsentation von Bedeutung ist ein Punkt, in dem viele Ansätze von Grundannahmen der Embodied Cognition abweichen. So finden sich in verschiedenen Ansätzen propositionale Strukturen oder andere explizite semantische Formalismen als Grundlage der Bedeutungsrepräsentation und der Konstruktion komplexer Bedeutungsstrukturen im Verstehensprozess. Propositionen sind ein klassisches Beispiel symbolischer Repräsentation: Sie sind abstrakte amodale Speicherstrukturen, die im Sinne eines Logikkalküls die Zusammenhänge atomarer Bedeutungskomponenten darstellen. Komplexe Ereignisse werden anhand diskreter atomarer Propositionen erfasst, die über algorithmische Mechanismen (Abrufprozesse und Interaktionen zwischen Langzeit- und Kurzzeitspeicher) in kompositionale propositionale Strukturen überführt werden. Damit widersprechen sie den Annahmen, die übereinstimmend in verschiedenen Ansätzen der Embodied Cognition getroffen werden (u.a. Barsalou, 1999; Zwaan, 2004). Propositionen finden sich unter anderem im CSCP-Modell (Rohde, 2002), in den SG-Modellen (McClelland et al., 1989; St. John, 1992), im Erwerbsmodell von Desai (2001, 2007), dem Ansatz von Dominey (2003); Dominey und Boucher (2005), dem hebbischen kompetitiven Netzwerk von Hadley et al. (2001) und im CIA-Net (Mayberry et al., 2009). Im INSOMNet (Mayberry, 2003) wird der Formalismus der Minimal Recursion Semantics verwendet, der als Erweiterung propositionaler Strukturen durch Verweismechanismen verstanden werden kann.

Nun ist für die anderen Modelle, die mit propositionalen Strukturen operieren, allerdings zu hinterfragen, inwiefern diese Input-Enkodierung ein Problem darstellt. So könnte man argumentieren, dass Propositionen in diesen Modellen nur als Input-Struktur verwendet werden, über die sich im Modell verteilte, nicht-propositionale Strukturen entwickeln sollen. Schließlich werden durch das Training mit dem propositionalen Input selbstorganisierend<sup>20</sup> komprimierte und verteilte Repräsentationen entwickelt. Sind diese internen Repräsentationen dann noch propositionaler Natur? Die tatsächlichen Verhältnisse bedingen wohl einen etwas genaueren Blick. Die internen Repräsentationen konnektionistischer Netzwerke entwickeln sich als emergente Struktur aus verschiedenen Interaktionen: (a) Interaktionen zwischen den Stimuli innerhalb einer Quelle, wobei sowohl die Strukturen und distributionalen Eigenschaften des Inputs, als auch im gegebenen Falle der überwachten oder selbstüberwachten Trainings der Target-Vektoren relevant sind; (b) Interaktionen zwischen den Stimuli aus verschiedenen Quellen sowie den verschiedenen Aufgaben betreffend der unterschiedlichen erforderlichen Transformationen für diese Quellen; (c) Interaktionen zwischen den Stimuli und dem System – seinen Eigenschaften wie etwa Kapazität, Lernrate, Lernalgorithmen, Rekurrenz, Kontext-Informationen, bisher entwickelte interne kategoriale Struktur etc. Dabei entwickelt das Netzwerk zwar eigenständig

---

<sup>20</sup>Die Modelle mit Abfrage-Mechanismus (z.B. Rohde, 2002; McClelland et al., 1989) werden zwar überwacht mit expliziten Target-Vektoren trainiert, die internen Repräsentationen der Sentence-, Story- oder Message-Gestalt entwickeln sich aber selbstorganisiert bzw. werden nicht explizit vorgegeben. In anderen Modellen (z.B. Mayberry, 2003; Hadley et al., 2001) entwickeln sich die Bedeutungsrepräsentationen über selbstorganisierte Trainingsalgorithmen wie zum Beispiel SOMs oder kompetitive Layer.

verteilte interne Repräsentationen, aber diese internen Repräsentationen bilden natürlich die Verhältnisse und Strukturen der Stimuli und der dazugehörigen Aufgaben ab. Denn schließlich müssen sie diese Stimuli und Aufgaben verarbeiten und bewältigen. Damit nimmt die vorgegebene Struktur der Stimuli erheblichen Einfluss auf die Gestaltung und Eigenschaften der internen Repräsentationen. Und damit handelt es sich unseres Erachtens bei den entsprechenden Modellen um konnektionistische Implementierungen propositionaler Modelle – oder zumindest um propositional induzierte Modelle.

Vor allem (S. L. Frank et al., 2009) äußern daran aus zwei Gründen explizit Kritik (vgl. Kap. 4.5): Erstens erfüllt ein Ansatz, der im Sinne der Embodied Cognition die Verwendung abstrakt-amodaler symbolischer Strukturen wie etwa Propositionen vermeidet, eine Grundvoraussetzung dafür, dass die Systematizität gefundener Ergebnisse auf einer genuin (eliminativ) konnektionistischen Lösung beruhen, und nicht nur auf die Implementierung symbolischer Strukturen und Prinzipien zurückzuführen sind. Abgesehen davon besteht die Gefahr, dass in den gewählten Repräsentationen semantische Intuitionen der Modellierer transportiert werden (vgl. Waskan, 2001). Diese führen das Modell auf genau die Lösungen hin, die den Modellierern plausibel erscheinen, und begünstigen genau die semantischen Dimensionen, die die Modellierer für entscheidend halten. Zweitens sprechen Befunde der Forschung in der Tradition des Embodied Language Processing für die unmittelbare Nutzung von Informationen über perzeptuelle und motorische Details sowie erfahrungsbasierte Stereotypizität und Gradualität (vgl. Kap. 8.3). Diese Informationen ermöglichen direkte Inferenzen, die in einem abstrakt-amodalen Repräsentationsformat nicht gegeben sind bzw. deren Realisierung in solchen Ansätzen schwer umzusetzen und zu begründen sind. Entsprechende Eigenschaften bleiben in Modellen auf propositionaler Basis ausgeblendet.

Der Kritik wäre noch ein dritter Grund bzw. eine Konkretisierung des ersten Einwands hinzuzufügen. Die Vorgabe propositionaler Strukturen nimmt dem Modell die Arbeit ab, die es selbst leisten soll – die Struktur und die Zusammenhänge in der Welt zu erkennen und in einer konzeptuellen Struktur abzubilden. Versteht man in den Modellen das propositionale Format als Gegenstand der mentalen Struktur, so müsste eigentlich erst erklärt werden, über welche Prinzipien, Mechanismen und Prozesse das Modell zu diesem Format gelangt. Versteht man hingegen das propositionale Format als Gegenstand der Umwelt, der das Modell ausgesetzt ist, so stellt man gänzlich falsche Voraussetzungen her – denn die Welt ist nicht offen propositional strukturiert, auch wenn man die Zusammenhänge in der Welt anhand von Propositionen darstellen kann. In der Folge dürfte es schwierig werden, zu identifizieren, welche Eigenschaften und Leistungsfähigkeiten durch das Modell aufgrund inhärenter Eigenschaften erworben wurden und welche Aspekte der Natur den vorgegebenen Stimuli zuzuweisen sind. Nur dies würde aber erlauben, sauber abzugrenzen, welche Erklärungsansätze für die menschliche Kognition aus der Datenstruktur abzuleiten sind und welche aus dem Modell.

Hinsichtlich des Repräsentationsformats ist auch das DSS-Modell (S. L. Frank et al., 2009, 2003b) nicht unproblematisch. Zwar stimmen wir mit der Einschätzung der Autoren überein, dass es sich bei der Enkodierung der Stimuli um ein nicht-propositionales Repräsentationsformat handelt. Doch die postulierte Analogizität zur abgebildeten Welt ist sehr abstrakt. Anhand von Korrelationen werden Konjunktion und Disjunktion von identifizierten und diskriminierten Ereignissen erfasst. Damit wird zum einen die Analogizität als sehr abstrakte Eigenschaft idealisiert. Zum anderen ist die geforderte Modalität der

Repräsentationen überhaupt nicht gegeben – ganz im Gegenteil wird sogar jede Modalität der Repräsentationen durch die Orthogonalisierung der lokalistischen Repräsentationen eliminiert (dies wird auch von den Autoren selbst klar gestellt, vgl. S. L. Frank et al., 2009; S. L. Frank, 2006a). Wir bezweifeln, dass ein solcher Ansatz, der die Wichtigkeit der Reduktion jeglicher expliziten Information über den Inhalt von Ereignissen und Objekten betont, mit der Annahme eines reichhaltigen multimodalen Repräsentationssystems kompatibel sein kann. Will man die Reichhaltigkeit der möglichen Strukturen und Zusammenhänge der Welt abbilden, bietet eine propositionale Form natürlich wesentliche Vorteile, um in sparsamer Weise möglichst viel Informationen darzustellen. Und zugegebenermaßen ist es schwierig, sich eine Semantikkonzeption vorzustellen, mit der auf vergleichbar einfache Art und Weise eine solche Bandbreite an semantischen Aspekten und Phänomenen modelliert werden kann, wie sie in den beschriebenen Modellen erfasst wird. Es stellt sich also die Frage, inwiefern die Implementierung einer alternativen nicht-propositionalen Konzeption des Sprachverstehens in einem konnektionistischen Modell überhaupt möglich bzw. sinnvoll umsetzbar ist. Wir werden darauf im folgenden Kapitel 10 wieder zu sprechen kommen, wenn es bei der Entwicklung einer geeigneten Repräsentation genau um dieses Problem gehen wird.

**Umsetzung des Wissenserwerbs** Ein weiteres Problem betrifft die Art und Weise, in der der Wissenserwerb in den verschiedenen Modellen umgesetzt wird. Dabei ist festzuhalten, dass in den meisten Modellierungen explizit darauf verwiesen wird, dass es sich nicht um Modelle des Spracherwerbs handelt. Die Form des Trainings soll also keine Erwerbsprozesse reflektieren, sondern ist zumeist rein pragmatisch der Generierung der internen Strukturen für das ausgereifte System geschuldet. Dies bedeutet, dass die Details des Erwerbs und der Verlauf der Performanzentwicklung nicht auf empirische Daten der Spracherwerbsforschung bezogen werden. Dies kann aber keine Rechtfertigung dafür sein, Mechanismen oder Trainingsabläufe für das Modell heranzuziehen, die jegliche Plausibilität des Erwerbs und der Erwerbssituation unberücksichtigt lassen.<sup>21</sup> Denn die Plausibilität der Art und Weise, über die das Modell zu seinem Wissen gelangt, erhöht auch die Plausibilität des Modells selbst. Wird diesem Aspekt widersprochen, so wird zumindest die Korrespondenz des Lernens (Simon & Wallach, 1999, vgl. Kap. 2.1) als Modellierungskriterium verletzt. Zumal sind konnektionistische Modelle im hier diskutierten Kontext Lernmodelle. Müssen für die Entwicklung der benötigten Wissensstrukturen unrealistische Mechanismen oder Verläufe angenommen werden, so stellt sich die Frage, ob diese Strukturen mit realistischen Mechanismen oder Verläufen überhaupt erworben werden könnten.

Zu diesem Gesichtspunkt sind verschiedene Kritikpunkte gegenüber den besprochenen Modellen zu äußern, die in erster Linie die Abbildung zwischen Sprache und Bedeutung betreffen. (a) Problematisch ist die in verschiedenen Modellen vorgenommene Trennung des Erwerbs von semantischem Wissen und Sprache: So werden im CSCP-Modell und im DSS-Modell erst die Repräsentationen der vollständigen semantischen Struktur bzw.

---

<sup>21</sup>Das bedeutet nicht, dass in jedem Modell jegliche Stufen des Erwerbsprozesses beginnend ab der ersten phonetischen Diskriminierung berücksichtigt werden müsste. Es ist absolut gerechtfertigt, einem Modell eine gewisse Abstraktionsebene und Erwerbsstufe zuzuschreiben und den Aspekt in den Vordergrund zu stellen, für den Erklärungen gesucht werden.

des vollständigen Situationsmodells entwickelt. Nach Abschluss dieses Erwerbs wird die Sprache auf diese Repräsentationen abgebildet. Dies ist unseres Erachtens nicht nur unrealistisch, sondern lässt auch Möglichkeiten der Erwerbssituation außer Betracht, die in einem synchronen holistisch-integrativen Training durch die gegenseitigen Beschränkungen verschiedener Informationsstrukturen und -kanäle gegeben sein könnten – wie etwa syntaktisches (Gleitman, 1990) oder semantisches (Pinker, 1984) Bootstrapping (vgl. auch Desai, 2007). (b) Die Frage der Informationsintegration betrifft auch die Umsetzung des Language Grounding. Dieser Aspekt wird in Modellen mit Dual-Route-Architektur oder mit mehreren Input-Modalitäten direkt umgesetzt. In allen anderen Architekturen bleibt dieser Aspekt hingegen ausgeblendet. Eine solche Umsetzung verhindert unter anderem die Umsetzung von Simulationen von situierter gegenüber entkoppelter Sprachverarbeitung, die für Fragestellungen der Embodied Cognition von Interesse sein könnten. (c) Des Weiteren ist die Umsetzung der Etablierung von Eins-zu-eins-Korrespondenzen zwischen sprachlicher Form und nicht-sprachlicher Bedeutung zu diskutieren. Alle genannten Modelle haben gemeinsam, dass sie die Abbildung von ganzen Sätzen mit mehreren Wörtern auf mehr oder weniger komplexe Szenen mit mehreren Objekten, Agenten, Ereignissen und Relationen leisten. Folglich müssen diese Modelle automatisch das Problem der Viele-zu-viele-Abbildung und der Etablierung von Eins-zu-eins-Korrespondenzen lösen, das in allen uns bekannten Modellen des lexikalischen Erwerbs ausgeblendet wurde (vgl. Kap. 9.3.1). Allerdings wird diese Aufgabe in den Modellen verschiedentlich vereinfacht: Im CSCP-Modell sowie im DSS-Modell muss der jeweilige Satz immer vollständig auf die semantische Repräsentation abgebildet werden. Es bleiben also keine semantischen Aspekte, die das System zu ignorieren lernen muss. Im CIA-Net muss zwar eine Auswahl aus zwei Handlungen getroffen werden. Allerdings wird die Abbildung des Satzes auf nur eine Handlung trainiert. Zudem wird die Auswahl der dafür relevanten Input-Informationen durch einen expliziten Aufmerksamkeits-Mechanismus unterstützt (siehe Diskussion folgender Abschnitt). Und in INSOMNet schließlich muss das Netzwerk zwar den für ein Input-Wort zuständigen Frame-Node bestimmen, auf diesen werden dann aber direkt die wortspezifischen und thematisch-rollebezogenen semantischen Informationen abgebildet. (d) Schließlich bleibt noch die Implementierung spezieller Mechanismen zu beurteilen. So wird im SG-Modell sowie im CSCP-Modell ein Abfrage-Mechanismus implementiert, dessen psychologische Plausibilität zumindest fragwürdig ist und in erster Linie der pragmatischen Umsetzung einer freien verteilten Enkodierung und Komprimierung der Stimuli zugeschrieben werden kann. Mit den Auffassungen von Erwerb, wie sie in Theorien der Embodied Cognition vertreten werden, hat dies jedenfalls nichts gemein.

**Explizite Verarbeitungsmechanismen** Abseits des Erwerbsaspekts wurde für CIA-Net ein spezieller Mechanismus entwickelt: Hier wird unbewusste Aufmerksamkeit (covert attention) durch einen Gating-Vektor als expliziter Aufmerksamkeits-Mechanismus umgesetzt. Dieser legt wie die propositionalen Repräsentationen den Verdacht der Implementierung eines klassisch-symbolischen Mechanismus nahe, der im Sinne eines Steuerungs-Algorithmus innerhalb des Systems Ressourcen verteilt und zuweist. Dies ist eher aus der Sicht des eliminativen Konnektionismus kritisch zu bewerten. In der Embodied Cognition hingegen spielt selektive Aufmerksamkeit zwar durchaus eine Rolle (vgl. Barsalou, 1999), wie diese jedoch im System umgesetzt ist, wird nicht spezifiziert. Kritik an einem externen expliziten Mechanismus äußern (Altmann & Mirkovic, 2009), da ein solcher erst durch

die Annahme der Unabhängigkeit der repräsentationalen Substrate etwa für sprachliche Bedeutung und szenisch-situationale Informationen und durch den Bedarf der Koindizierung dieser Substrate erforderlich wird. In ihrer Konzeption stellen sie dem expliziten einen impliziten Mechanismus gegenüber, der dadurch zustande kommt, dass sich sowohl Aufmerksamkeit als auch sprachliche und nicht-sprachliche Informationen dasselbe repräsentationale Substrat teilen. Aufmerksamkeit wird implizit durch verschiedene Zustände dieses Substrats kontrolliert.

Hinsichtlich externer Mechanismen wäre noch das Modell des Erwerbs grammatischer Konstruktionen von Dominey und Boucher (2005) zu erwähnen.<sup>22</sup> Dieses ist als Roboter-Modell mit konnektionistischem Kernmodul implementiert und besitzt durch seine hybride Architektur echte modale Kanäle. Auf dieser Grundlage simuliert es beschränkt eine echte Interaktion mit einem Gesprächspartner. Allerdings ist die Architektur stark modular aufgebaut und übersetzt die modalen Informationen – von Informationsintegration und analog-modaler Repräsentation kann hier keine Rede sein.<sup>23</sup> Zudem verwendet auch dieses Modell propositionale Target-Repräsentationen und noch dazu statische Input-Repräsentationen der Sätze. Gerade die Mechanismen, die diese Strukturen erzeugen, sind nicht konnektionistisch realisiert. Damit bleibt nur ein recht bescheidener Anteil der Leistungsfähigkeit, dessen Eigenschaften durch konnektionistische Prinzipien erklärt werden können.

### 9.3.3. Kriterien für Modelle des Satzverstehens

Wir haben im vorausgegangenen Unterkapitel verschiedene Aspekte benannt, hinsichtlich derer konnektionistische Modelle des Satzverstehens als Implementierungen der Sprachverarbeitung im Embodiment-Ansatz verstanden werden können. Dem wurden andere Aspekte entgegengestellt, hinsichtlich derer diese Modelle von Annahmen des Embodiment-Ansatzes fundamental abweichen. Es bleiben folgende Fragen zu beantworten: Wie müsste ein konnektionistisches Modell aussehen, das diese Annahmen modelliert? Welche Kriterien müsste es erfüllen und welche Aspekte berücksichtigen? Auf Grundlage der im vorausgehenden Kapitel 8.2 und 8.1.2 diskutierten Theorien zur Sprachverarbeitung in der Embodied Cognition und mit Blick auf die besprochenen Modelle schlagen wir einige Kriterien vor.

**Analogizität** Das konzeptuelle Wissen sollte die Analogizität zum repräsentierten Weltausschnitt erhalten. Folglich sollten auch die Enkodierungen der Stimuli diese Analogizität zumindest in den für die Modellierung relevanten Dimensionen bewahren. Umgekehrt sollten mit den Stimuli keine Informationen enkodiert sein, die eine andere als die in der Welt tatsächlich wahrnehmbare Struktur induzieren und damit das Verhältnis zwischen internen Strukturen und den Strukturen und Zusammenhängen in der Welt verzerren würden.

---

<sup>22</sup>Das Modell von Dominey und Boucher (2005) wird im Zusammenhang zu verschiedenen Ansätzen in der Embodiment-Tradition gesehen, zum Beispiel verschiedenen Ansätzen des L<sub>0</sub>-Projekts oder den Roboter-Modellen von Steels et al. (2002).

<sup>23</sup>Diese Kritik kann wohl gegenüber allen gegenwärtigen Roboter-Ansätzen geäußert werden.

**Multimodalität** Die Enkodierung der Stimuli sollte sich an Informationsstrukturen menschlicher Modalitäten bzw. an den Repräsentationsformaten sensomotorischer Interaktion orientieren, um direkte Inferenzen zu ermöglichen. Mit dem Input sollten also Informationen dargeboten werden, die zum Beispiel dem visuellen, auditorischen, haptischen oder olfaktorischen Kanal oder der motorischen Kontrolle des körperlichen Bewegungsapparats entsprechen.<sup>24</sup> Dabei sollten die Stimuli deren Eigenschaften bewahren, aber nicht um zusätzliche Informationen erweitern, die explizite semantische Strukturen vorgeben.

**Informationsintegration** Die Architektur des Modells sollte die Möglichkeit eröffnen, Informationen verschiedener Modalitäten oder Verarbeitungsdomänen zu integrieren und Interaktionen verschiedener Informationen und Aufgaben bei der Verarbeitung und internen Strukturierung zu ermöglichen.<sup>25</sup>

**Situationalität** Das Modell sollte die Situationalität von Erwerbs- und Verarbeitungssituationen erfassen. Zugleich sollte sie aber auch die Simulation von Verarbeitungsprozessen ermöglichen, die von der aktuellen Wahrnehmung oder Situation entkoppelt sind.

**Emergenz** Die für die Verarbeitung relevanten Wissensstrukturen und Mechanismen (z.B. Aufmerksamkeit, Kategorisierung, Schematisierung, Zuweisung) sollten sich selbstorganisierend aus der Interaktion von System und Umwelt entwickeln. Die Induzierung entsprechender Strukturen durch explizite semantische Informationen oder durch explizit implementierte Mechanismen (etwa Selektions- oder Aufmerksamkeitsmechanismen) sollte vermieden werden.<sup>26</sup>

**Generalisierung** Das Modell sollte sich nicht auf auswendig gelernte Strukturen oder starre Assoziationen holistischer Repräsentationen reduzieren lassen. Es sollte Effekte der Selektion relevanter Informationsaspekte, der Schematisierung und der Kategorisierung der Stimuli aufweisen. Zudem sollte es ein gewisses Maß an Systematizität zeigen, das das Verstehen noch nicht in dieser Form gehörter Sätze und die Konstruktion noch nicht in dieser Form gesehener Situationsmodelle ermöglicht.

Ist die Forderung der Einhaltung all dieser Vorgaben und zugleich die Simulation der Verarbeitung einer Sprache von hinreichender Komplexität realistisch? Ist die Einhaltung aller Vorgaben überhaupt erstrebenswert? Dies hängt wohl von den Zielen der Modellierung ab. Zwar wäre ein vollständiges Modell innerhalb gewisser Grenzen denkbar, allerdings würden wir dabei sicherlich sehr schnell bei einer Large Scale-Modellierung mit einer unüberschaubar komplexen Architektur enden. Wenn wir den Konnektionismus

---

<sup>24</sup>Dies setzt natürlich voraus, dass wir erstens die Informationsstruktur kennen (was etwa bei neuronalen Repräsentationen motorischer Kontrollstrukturen eher schwierig sein dürfte) und zweitens, dass wir diese Informationsstruktur überhaupt in einer entsprechenden Vektor-Kodierung für die Modellierung umsetzen können.

<sup>25</sup>Bedeutung sollte also zeitgleich und in Interaktion mit sprachlicher Form erworben werden, um damit den situierten Spracherwerb zu simulieren. Die Methode des getrennten Trainings für verschiedene Module und Input-Domänen würde diesem Kriterium widersprechen.

<sup>26</sup>Die Verwendung jeder expliziten Enkodierung von Semantik, die sich nicht durch Modalität oder Analogizität begründen lässt, muss als potenzielle Einschränkung der Aussagekraft für die Embodiment-Modellierung beurteilt werden. Dies ist nicht als Kritik an entsprechenden Modellen selbst zu verstehen, sondern als Einschränkung ihrer Aussagekraft für Aspekte der Embodied Cognition.

aber als Plattform für Fragestellungen der Embodied Cognition nutzen wollen oder spezielle theoretische Annahmen spezifizieren wollen, ist es unseres Erachtens sinnvoller, die Forderungen zu beschränken und nicht eingehaltene Forderungen aus dem Erklärungsspektrum des Modells auszuschließen. Denn vielleicht können umfassende Modelle, die alle Dimensionen berücksichtigen, eine validere Performanz beanspruchen. Zugleich verlieren sie aber auch das, was Pospeschill (2004) als grundlegendes Kriterium der Erklärungsmächtigkeit von konnektionistischen Modellen fordert – ihre Transparenz. Large Scale-Modelle sind berechtigt und wichtig für die konnektionistische Forschung, vor allem in Anbetracht der Infragestellung der Skalierbarkeit von Netzwerkmodellen auf ernstzunehmend komplexe Gegenstandsbereiche. Allerdings werden sie durch ihre Komplexität zu Black-Box-Modellen, die die dahinterstehenden Prozesse und Mechanismen unzugänglich machen.<sup>27</sup> Kello und Plaut (2003) unterschieden in diesem Zusammenhang realistische und fundamentalistische Modellierungsansätze: Erstere versuchen, möglichst viele Details und damit jegliche Interaktion zu erfassen, die möglicherweise für die zu erklärenden Phänomene relevant sein könnte. Zweitere versuchen, nur die Details zu erfassen, die für die zu erklärenden Phänomene tatsächlich als relevant erachtet werden (vgl. auch Plaut, 2002). Als effektiver Modellierungsansatz wird dabei erachtet, erst anhand fundamentalistischer Ansätze die zentralen Faktoren und Prinzipien zu identifizieren. Dann wird schrittweise die Komplexität des Modells erhöht, um die Auswirkung zusätzlicher Details zu untersuchen. Dies scheint uns eine plausible und pragmatische Herangehensweise zu sein.

---

<sup>27</sup>Zudem steigt mit zunehmender Komplexität auch die Anzahl manipulierbarer Parameter, wodurch ein beliebiges ‚Tuning‘ des Netzwerkverhaltens begünstigt wird.

## 10. Konnektionistische Modellierung: Sprachemergenz und Embodiment

Wie sind unsere Sprache, unser Denken, unser Körper und die Welt, auf die wir uns beziehen, ineinander verankert? Und wie muss unser kognitives System beschaffen sein, damit es diese Verankerung herstellen kann? Wir haben in den vorausgegangenen Kapiteln 8 und 9 theoretische Ansätze, empirische Studien und Modellierungen besprochen, die zu zeigen versuchen, wie unsere mentalen konzeptuellen Strukturen und unsere Sprache in der Außenwelt und unseren körperspezifischen Wahrnehmungs- und Interaktionsmöglichkeiten gegroundet sein könnten. Im folgenden Kapitel präsentieren wir die Ergebnisse des Modellierungsprojekts *SynSem* (der Arbeitstitel des Projekts steht für Syntaktisch-Semantisches Integrationsmodell). Im Zentrum des Projekts steht ein konnektionistisches Modell, das das Grounding von Sprache und (pseudo-)visueller Verarbeitung simuliert. Gegenstände der Modellierung sind die Verarbeitung sequenziell dargebotener Sätze, das Verstehen dieser Sätze und der dafür erforderliche Bedeutungserwerb. Die Sätze beziehen sich auf kleine statische visuelle Szenen von Objekten in einer räumlichen Anordnung, die Verstehensdomäne ist also die räumliche Semantik. Das Modell simuliert, wie im Bedeutungserwerb die Viele-zu-viele-Korrespondenz zwischen Mehrwort-Äußerungen und visuellen Szenen mit mehreren möglichen Bedeutungsaspekten als situationsübergreifendes Lernen und Bootstrapping-Prozess erfolgen kann. Sprachverstehen wird als Indizierung von sprachlicher Form auf nicht-sprachliche modale Bedeutung erworben. Das Modell ermöglicht dabei verschiedene Modi der Sprachverarbeitung: entkoppeltes (d.h. situationsunabhängiges) Sprachverstehen, situiertes Sprachverstehen sowie in beschränktem Umfang Sprachproduktion. Wie zeigen, wie die dafür erforderlichen Leistungen als emergente Produkte der Interaktion verschiedener Informationen und Transformationsaufgaben entstehen und wie sich dabei im Sinne von Constraint-Satisfaction-Prozessen erweiterte kognitive Leistungen wie Inferenzen, Kontrolle der visuellen Aufmerksamkeit und sprachliches Priming entwickeln.

Wir stellen zunächst unseren Modellierungsansatz dar und erläutern die Architektur und die Gestaltung der Stimuli. Dieser Ansatz orientiert sich an den Kriterien, die wir in Kapitel 9.3.3 vorgestellt haben. Anschließend präsentieren wir zwei größere Simulationsstudien, die auf dieser Grundlage durchgeführt wurden. In *SynSem I* demonstrieren wir die prinzipielle Leistungsfähigkeit des Modells. In *SynSem II* heben wir einige Beschränkungen der ersten Studie auf und untersuchen einige erweiterte Fragestellungen. Wir besprechen die beiden Simulationen getrennt voneinander und erläutern für beide die zugrundeliegenden Fragestellungen, Hypothesen, Details der Modellierung und der Simulation und die durchgeführten Untersuchungen. Anschließend diskutieren wir die Ergebnisse der jeweils dazu durchgeführten Analysen für die verschiedenen Verarbeitungsmodi und erläutern sie anhand repräsentativer Beispiele der Netzwerkperformanz. Abschließend diskutieren wir

den Modellierungsansatz und die Ergebnisse in Beziehung zu den besprochenen theoretischen Ansätzen der Embodied Cognition und zu den Modellierungsansätzen zu Language Grounding und Embodied Language Processing. Das Modell und Ergebnisse aus den Simulationen wurden von Weldle, Konieczny et al. (2009a, 2009b); Weldle, Konieczny, Müller et al. (2010a, 2010b) dargestellt.

### 10.1. Modellierungsansatz des Projekts SynSem

In unserem Modellierungsansatz greifen wir verschiedene Ansätze auf, die wir zur konnektionistischen Modellierung des Sprachverstehens, des Language Grounding und des Embodiment besprochen haben. Diese Ansätze bestimmen auf die eine oder andere Weise (a) die gewählte Architektur, (b) die Gestaltung der Stimuli, (c) die Gestaltung des Trainingsablaufs, (d) die untersuchten theoretischen Fragestellungen, (e) die untersuchten Performanzaspekte der Netzwerke und (f) unsere daraus entwickelte Konzeption von Verstehen und Bedeutung. Wir stellen zunächst die Architektur vor und erläutern dann die verwendeten Stimuli und die Wahl der untersuchten Domäne. Anschließend wenden wir uns in Verbindung von Architektur und Stimuli einer Analyse der Aufgaben und der erforderlichen Leistungen zu, die das Netzwerk erbringen soll. Bezogen auf Architektur, Stimuli und Aufgaben diskutieren wir die Grundideen unseres Ansatzes und die Umsetzung der in Kapitel 9.3.3 angedachten Kriterien.

#### 10.1.1. Architektur und Aufgaben

Unser Ausgangspunkt ist eine konnektionistische Architektur, die eine direkte Abbildung von Sprache auf modale Wahrnehmungen von partiell zur Äußerung korrespondierenden Situationen vorsieht und damit einen situational verankerten Bedeutungserwerb vollzieht. Als Modalität haben wir die visuelle Verarbeitung räumlicher Situationen gewählt, die bereits mehrfach Gegenstand entsprechender Simulationen wurde (v.a. in hybriden und strukturiert konnektionistischen Modellen, etwa Regier, 1995; Sales et al., 1996; Feldman et al., 1996; Steels et al., 2002; Dominey & Boucher, 2005). Wir verwenden dabei eine primitive Mikrowelt, in der vier unterscheidbare Objekte in einem spezifischen räumlichen Verhältnis zueinander stehen können. Für die sprachlichen Äußerungen verwenden wir die Verarbeitung von Sätzen, die sequenziell Wort für Wort dargeboten werden. Diese sprachlichen Äußerungen beschreiben die visuellen Situationen. Die Architektur muss bei Erwerb und Verarbeitung die Abbildung von inkrementeller und damit dynamischer Sprachverarbeitung auf statische Derivate der visuellen Verarbeitung leisten. Dies entspricht der Indizierung erfahrungsbasierter nicht-sprachlicher Konzepte durch sprachliche Formen, die in der Embodied Cognition weitgehend einheitlich als Grundlage des Sprachverstehens vorgeschlagen wurde (etwa Barsalou, 1999; Glenberg & Robertson, 2000; Zwaan, 2004; Bergen & Wheeler, 2009). Durch eine nur partielle Überlappung von Äußerung und Situation und die Verwendung von Mehrwort-Äußerungen muss dabei das Problem der Identifizierung korrekter Form-Bedeutungs-Abbildungen bei referenzieller Uneindeutigkeit (Quine, 1960) durch potenzielle Viele-zu-viele-Korrespondenzen gelöst werden (Davis,

2002). Dies ist ein grober Abriss des primären Leistungsspektrums, das durch das Modell abgedeckt werden soll.

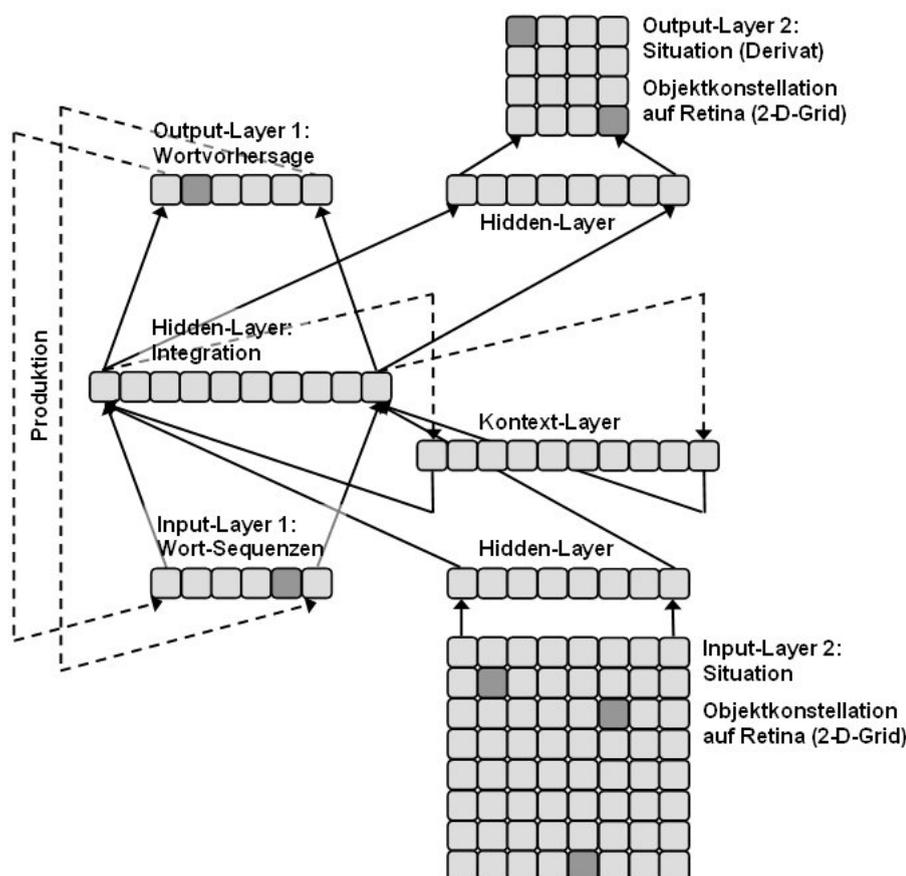


Abbildung 10.1.: Basis-Architektur des SynSem-Modells: Grundlage ist ein SRN, das mit zwei Input-Output-Routen versehen ist, die über den rekurrenten Hidden-Layer integriert werden. Die visuelle Route (rechte Seite) verarbeitet räumlich-relational verteilte Objekte, die auf einer zweidimensionalen Retina dargeboten werden. Die sprachliche Route (linke Seite) verarbeitet Sätze, die Wort für Wort dargeboten werden.

Um dieses Spektrum zu erfassen, verbinden wir die beiden vermutlich erfolgreichsten konnektionistischen Ansätze des Language Grounding und der sequenziellen Verarbeitung: Als Modell für das Grounding und die Integration der Informationen verwenden wir die Dual-Route-Architektur (Plunkett et al., 1992; Cangelosi, 2005), als Modell für die sequenzielle Verarbeitung der sprachlichen Äußerung verwenden wir die SRN-Architektur (Elman, 1990). Die Architekturen werden in einem SRN mit zwei Input-Output-Routen integriert, die die Verarbeitung verschiedener Domänen – verschiedener Informationsstrukturen, verschiedener hierarchische Abhängigkeiten, verschiedener Transformations-Aufgaben und einer unterschiedlichen temporalen Struktur – leisten (vgl. Abb. 10.1):

- Die visuelle Route verarbeitet Informationen aus einem zweidimensionalen Gitter von Units, einer Retina nachempfunden. Auf diesem werden verschiedene im Raum verteilte einfache Objekte dargestellt. Die zu leistende Transformations-Aufgabe gehört in den Bereich der Muster-Assoziation (SynSem I) bzw. der Auto-Assoziation (SynSem II).

- Die sprachliche Route verarbeitet Sätze, die Wort für Wort dargeboten werden. Die zu leistende Transformations-Aufgabe ist die klassische Wortvorhersage-Aufgabe für das jeweils folgende Wort der Sequenz.
- Zwischen den Routen herrscht eine Korrespondenz-Beziehung: Die Sätze korrespondieren partiell zum visuellen Input – sie beschreiben einen Teil der visuell dargebotenen Objekte und der räumlichen Relationen zwischen diesen.
- Das SRN integriert die beiden Routen über den Integrations-Layer, den gemeinsamen rekurrenten Hidden-Layer.

**Aktivierung und Fehlerpropagierung** Auf der sprachlichen Route wird der Input über eine vollständige Verknüpfung vom Input-Layer in den rekurrenten Integrations-Layer eingespeist. Dieser erhält zudem über eine vollständige Verknüpfung Informationen über den bisherigen sequenziellen Kontext aus dem vorhergehenden Zeitschritt aus dem Kontext-Layer. Der Integrations-Layer legt eine Kopie seiner Aktivierungs-Informationen des aktuellen Zeitschritts über eine Eins-zu-eins-Verknüpfung (mit auf den Faktor 1.0 fixierten Gewichten) auf dem Kontext-Layer ab. Über eine vollständige Verknüpfung propagiert er die Aktivierung schließlich zum Output-Layer der sprachlichen Route. Der Kontext-Layer erhält für die Erfüllung der Kopierfunktion eine lineare Aktivierungsfunktion, alle anderen Layer erhalten eine sigmoide Aktivierungsfunktion. Auf dem Output-Layer wird die Aktivierung mit einem Target-Vektor abgeglichen, der das jeweils nächste in der Wort-Sequenz folgende Wort angibt. Das Netzwerk soll also auf dieser Route lernen, das jeweils folgende Wort vorherzusagen. Der Fehler zwischen dem Aktivierungs-Vektor und dem Target-Vektor wird über die Kreuzentropie (cross entropy) der Vektoren bestimmt und über den Algorithmus Backpropagation Through Time (BPTT) über die internen Layer zum Input-Layer zurückpropagiert. Die Gewichte werden dementsprechend angepasst.

Auf der visuellen Route werden die Stimuli auf dem Input-Layer durch die Aktivierung einzelner Units als räumlich verteilte Punkt-Muster dargestellt. In der ursprünglichen Version der Architektur, die für SynSem I verwendet wurde, existiert keine Zweidimensionalität, wie sie durch die gegebene Abbildung oder die Präsentation der Stimuli suggeriert wird. Zwar wurde eine zweidimensionale räumliche Struktur des Layers für die Gestaltung der Stimuli zugrundegelegt. Allerdings werden die Stimuli anhand eines linearen Vektors dargeboten: Die Units des Gitters werden von links oben nach rechts unten in einen linearen Vektor überführt. Vom Input-Layer wird die Aktivierung über einen Kompressions-Layer über vollständige Verknüpfungen zum Integrations-Layer weitergeleitet. Der Kompressions-Layer ist kleiner als der Input-Layer und dient der Verdichtung der Informationen durch die Reduzierung der Dimensionen. Wie bereits für die sprachliche Komponente beschrieben, wird der aktuelle Inhalt des Integrations-Layers auf den Kontext-Layer kopiert und der Inhalt des vorausgehenden Zeitschritts zum aktuellen Input des Integrations-Layers hinzugefügt. Von dort wird die Aktivierung auf der visuellen Route weiter über einen zweiten Kompressions-Layer über vollständige Verknüpfungen zum visuellen Output-Layer propagiert. Dort erfolgt ein Abgleich mit einem Target-Vektor, der in einem späteren Abschnitt erläutert wird. Der Fehler wird wiederum über die Kreuzentropie (cross entropy) der Vektoren bestimmt. Auch die Propagierung des Fehlers und die Anpassung der Gewichte erfolgt wiederum über den Algorithmus Backpropagation Through Time (BPTT) über

alle internen Layer zurück zum Input-Layer der visuellen Route. Allerdings lässt sich der zurückpropagierende Fehler ab dem gemeinsamen Integrations-Layer nicht mehr klar für die beiden Informationsquellen trennen, denn hier wird denselben Fehleranteil aus zwei verschiedenen Fehlerquellen zugewiesen. Dies wird Konsequenzen für die entwickelten internen Repräsentationen haben, die wir mit der Rolle des Integrations-Layers besprechen.

**Variation der Architektur** Die hier beschriebene Architektur stellt die Basis-Architektur unseres Ansatzes dar. Diese wurde in verschiedener Weise modifiziert: Für SynSem I wurden zur Reduzierung der Informationen der visuellen Route und zur Unterstützung der Generalisierung die verschiedenen Kompressions-Layer eingefügt. Diese wurden (a) nur zwischen den visuellen Input und den Integrations-Layer, (b) nur zwischen den Integrations-Layer und den visuellen Output, (c) wie oben dargestellt auf beiden Positionen oder (d) in einer Version ohne Kompressions-Layer auf keiner der beiden Positionen geschaltet. Da diese Modifikationen keine qualitativen und nur geringe quantitative Unterschiede bei der Performanz bedingen, werden wir nur die Simulation zur Variante mit beiden Kompressions-Layern vorstellen, da sie die besten quantitativen Ergebnisse, d.h. die beste übergreifende Fehlerreduzierung, erbrachte.

Für SynSem II wurden einige weiter reichende Veränderungen vorgenommen, von denen sich aber nur eine dem von uns erwarteten Verhalten nähert. Die zwei wesentlichen Änderungen der erfolgreichen Variante betreffen den ersten Kompressions-Layer, der mit einer Funktion rezeptiver Felder ausgestattet wurde, sowie den visuellen Output-Layer, der im Sinne einer auto-assoziativen Abbildung hinsichtlich Größe und Target-Vektoren identisch zum visuellen Input-Layer ist. Einen kurzen Überblick über getestete Varianten und eine Erläuterung der erfolgreichen Modifikation verschieben wir auf den Bericht zur zweiten Simulation in Kapitel 10.3, wo wir auch die Hintergründe dazu erklären können.

Ein Modell zur Integration visueller und sequenziell-sprachlicher Verarbeitung auf Basis eines Dual-Route-SRN in der hier vorgestellten Form wurde unseres Wissens bislang nicht implementiert. In verschiedenen Modellen wird eine Anwendung normaler Dual-Route-Netzwerke auf die lexikalische Verarbeitung beschrieben (etwa Plunkett et al., 1992; Coventry et al., 2005). Allerdings wird hier die Aufgabe des Vokabelerwerbs stark vereinfacht, da verschiedene Abbildungsprobleme (etwa das Problem der Viele-zu-viele-Korrespondenzen) ausgeblendet werden (vgl. Kap. 9.3.1). In einer zu unserem Ansatz ähnlichen Architektur mit Dual-Route-SRN erläutern Dienes et al. (1999) die Abbildung verschiedener sequenzieller Struktursysteme aufeinander (z.B. für das Artificial Grammar Learning). Auf diesem Modell basierend entwickeln Altmann und Mirkovic (2009) eine Rahmentheorie des Sprachverstehens. In dieser werden die Eigenschaften von SRNs als allgemeine Modelle menschlicher Sprachverarbeitung diskutiert. Die Autoren erläutern die Rahmentheorie anhand theoretischer Anwendungsmöglichkeiten von SRNs für sprachliche und situationale Vorhersage-Aufgaben, und untermauern die angenommenen Eigenschaften mit empirischen Befunden (vgl. Kap. 4.4 und Kap. 9.3.2.1). Allerdings werden keine tatsächlichen Modellierungen umgesetzt, die Anwendungen bleiben spekulativer Natur.

### 10.1.2. Gestaltung der Stimuli

Bevor wir weitere Details zum Modell, zur Funktionsweise und zu den Erwartungen an die Architektur näher beleuchten, wenden wir uns kurz der Gestaltung der verwendeten Stimuli zu. Entsprechend den Annahmen des Embodied-Cognition-Ansatzes wird semantische Verarbeitung bzw. Verstehen als Abbildung von sprachlicher Form auf die erfahrungsbasierten Derivate der modalen Wahrnehmung von Situationen in der Welt konzipiert.

#### 10.1.2.1. Visuelle Stimuli

Die modale Verarbeitung von Situationen in der Welt wurde hier umgesetzt als Verarbeitung eines visuellen Stimulus auf einem retinaartigen zweidimensionalen Unit-Gitter. Das soll allerdings nicht bedeuten, dass es unsere Intention wäre, damit ein Modell visueller Verarbeitung zu entwickeln. Dies wird deutlich, wenn wir die Konstruktion der visuellen Stimuli und die zu erfüllenden Verarbeitungsleistungen des Modells betrachten (vgl. Abb. 10.2).

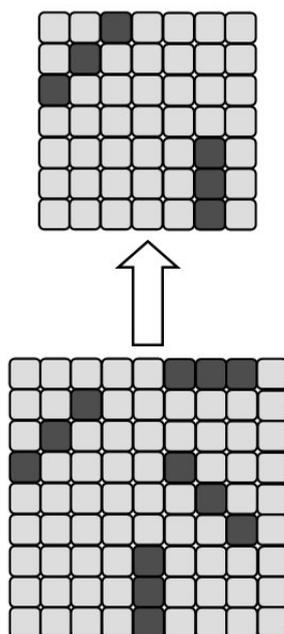


Abbildung 10.2.: Visuelle Stimuli der Mikrowelt für die Simulation SynSem I. Aufgabe der Transformation der beliebig angeordneten Objekte auf eine reduzierte und prototypisierte Version.

In unserer Mikrowelt lassen sich vier Objekte unterscheiden: Das MINUS, die PIPE, der SLASH und der BACKSLASH. Die Objekte werden durch die Aktivierung von jeweils drei einzelnen in der räumlichen Darstellung zusammenhängenden Punkten dargestellt. Von diesen Objekten können sich jeweils eines, zwei oder drei Objekte gleichzeitig in einer Situation befinden (niemals aber zwei gleiche Objekte). Die Objekte befinden sich dabei in einer beliebigen räumlichen Verteilung auf dem zweidimensionalen Grid, wobei sich die Objekte berühren, aber nicht überlappen können. In den sprachlichen Beschreibungen

werden acht räumliche Relationen zwischen den Objekten unterschieden: ÜBER, UNTER, RECHTSVON, LINKSVON, RECHTSÜBER, LINKSÜBER, RECHTSUNTER, LINKSUNTER. In unserem Beispiel 10.2 etwa befindet sich SLASH in den Relationen LINKSUNTER MINUS, LINKSÜBER PIPE sowie LINKSÜBER BACKSLASH.

Die Transformations-Aufgabe für die visuelle Route ist in Simulation SynSem I die Muster-Assoziation. Für jede dargebotene Situation wird eine reduzierte und idealisierte Version der Situation als Target dargeboten, die drei Aspekte umfasst: Die Reduzierung betrifft (a) die Größe des zweidimensionalen Gitters und folglich des abgebildeten Raums, (b) die Anzahl der Objekte – der Target-Vektor enthält zwei Target-Objekte und im Falle nur eines dargestellten Objekts entsprechend ein Target-Objekt. Die Target-Objekte sind eine Auswahl der als Input dargestellten Objekte. Die Idealisierung betrifft (c) die Positionen der Objekte, die die Positionierung der entsprechenden Input-Objekte in idealisierter Form abbildet. In Beispiel 10.2 etwa wird aus den dargebotenen Objekten der SLASH und die PIPE ausgewählt. Ihre relative Positionierung wird auf die idealisierte LINKSÜBER- bzw. RECHTSUNTER-Relation abgebildet. Diese Reduzierung und Idealisierung wurde handkodiert als Target-Aktivierung vorgegeben, wobei die Bestimmung der idealisierten Relation anhand einer Winkelfunktion aus dem jeweiligen Input-Muster berechnet wurde.<sup>1</sup> Dies ist ein nicht unproblematischer Sachverhalt, denn zugegebenermaßen ist die Darbietung solcher Target-Repräsentationen eine Form expliziter Struktur- und Informationsvorgabe. Denn sie übernimmt mehrere Aufgaben, unter anderem gibt sie sozusagen den Aufmerksamkeitsfokus auf bestimmte Objekte vor und bestimmt die Prototypisierung der räumlichen Relationen. Gerade die Vorgabe expliziter Informationen wollen wir in unserem Ansatz aber vermeiden (vgl. Diskussion im folgenden Absatz). Deshalb hat dieser Sachverhalt Konsequenzen für die Gestaltung der Transformations-Aufgabe für die Simulation SynSem II. Auf die einzelnen Kritikpunkte und die Neugestaltung der Aufgabe werden wir an späterer Stelle zu sprechen kommen (vgl. Kap. 10.3).

**Begründung der Wahl des Repräsentationsformats** Bereits an der Reduziertheit unseres Ansatzes zur Verarbeitung einer visuellen Szene dürfte erkennbar sein, dass es uns nicht um ein Modell visueller Verarbeitung gehen kann. Ein ernsthaftes kognitives Modell visueller Verarbeitung würde wesentlich umfangreicheren Ansprüchen genügen müssen als denen, die unser Modell erfüllen kann.<sup>2</sup> Vielmehr leistet der gewählte Ansatz eine Annäherung daran, die nur einem Ausschnitt der erforderlichen Performanzaspekte ge-

<sup>1</sup>Die Erstellung der Stimuli mit allen Kombinationen in beliebiger Auswahl sowie mit Berechnung der idealisierten Target-Aktivierung erfolgte anhand eines in CommonLisp programmierten Tools. Das Tool für SynSem I wurde von Daniel Müller erstellt. Hierfür an dieser Stelle herzlichen Dank.

<sup>2</sup>So sind einerseits die visuellen Stimuli stark reduziert: Die unterscheidbaren Objekte verfügen über ein Minimum an Komplexität, es werden keine Farben und keine Lichtverhältnisse unterschieden, die Komplexität lässt keine weitergehende Unterscheidung von spezifischen Eigenschaften zu, die Objekte sind nicht-beweglich und interagieren nicht miteinander, der Raum und die Objekte sind nicht dreidimensional (es muss also keine Projektion auf eine zweidimensionale Repräsentation und auch keine mentale Rückübersetzung in einen dreidimensionalen Raum stattfinden), und es gibt keine Vordergrund- oder Hintergrund-Objekte. Physiologische Aspekte der visuellen Wahrnehmung, die etwa durch die motorischen und optischen Eigenschaften der Augen vorgegeben sind, werden nicht berücksichtigt. Dies gilt auch für die neuropsychologischen Aspekte, die die neuronale Architektur der visuellen Verarbeitung betreffen. So werden verschiedene spezialisierte Subprozesse und Pfade der Verarbeitung nicht differenziert – nicht einmal die elementare Trennung der Pfade für die Verarbeitung von Objekten

recht wird. Weshalb aber haben wir uns dann für eine quasi-retinale Schnittstelle und ein pseudo-visuelles Repräsentationsformat entschieden?

Für die Auswahl dieses Repräsentationsformats gibt es eine Verkettung aufeinander bezogener Gründe: Erstens besitzt es weit mehr Eigenschaften visueller Perzeption, als dies bei den anderen aktuellen Modellen konnektionistischen Sprachverstehens der Fall ist. In allen anderen Ansätzen werden die wahrgenommenen Objekte und Eigenschaften direkt durch einzelne Units oder Vektor-Ausschnitte bezeichnet und die Relationen der Bestandteile der visuellen Situation werden explizit vorgegeben (vgl. z.B. Mayberry et al., 2009). Oder aber es wird über alle perzeptuellen Merkmale hinweg auf Gesamtsituationen abstrahiert (z.B. S. L. Frank et al., 2009). Im Vergleich dazu besitzt das von uns gewählte Format eine wesentlich größere Nähe zu tatsächlicher modaler Wahrnehmung und damit das Potenzial, analog-modale interne Repräsentationen zu induzieren. Zweitens wird in diesem Format für jede Input- und Target-Unit nur bestimmt, ob ein perzeptueller Reiz zugegen ist oder nicht. Der semantische Gehalt, der den einzelnen Units dabei zugewiesen wird, ist also auf ein absolutes Minimum reduziert. Die Aktivierung eines Knotens verweist nicht auf explizite semantische Informationen oder semantische Relationen. Zudem sind die Knoten nicht in semantisch bedeutsame Einheiten gruppiert, über die den Bedeutungseinheiten im Sinne eines propositionalen Formats Argumente und Relationen zugewiesen werden könnten. Eine beliebige Unit hat nichts und alles mit den unterscheidbaren Objekten und räumlichen Relationen zu tun, und das lokale Muster eines Objekts verweist unabhängig von seiner Position auf dem Input-Vektor auf dieses Objekt. Mit dieser Form der Reduzierung wählen wir für die Repräsentation sozusagen den genau entgegengesetzten Ansatz zu dem, der mit dem DSS-Modell gewählt wird (vgl. S. L. Frank et al., 2003b, 2008). Gerade dieser Unterschied erlaubt es uns aber, außer der Analogizität die Modalität der daraus abgeleiteten internen Repräsentationen zu gewährleisten und in den Grenzen unserer Mikrowelt entsprechende modale Inferenzen zu erfassen. Drittens ermöglicht uns die Orientierung an unstrukturierter visueller Perzeption, ein Repräsentationsformat zugrunde zu legen, dessen Struktur und Informationsgehalt möglichst wenig durch uns beeinflusst werden kann. Durch diese absolute Selbstbeschränkung wollen wir erreichen, dass keine explizit strukturierte Information über semantische Relationen gegeben wird. Denn diese soll das Netzwerk schließlich selbstständig erkennen und erlernen. Unser Ziel ist es, dass das Netzwerk die Leistungen der Extraktion und der Zuweisung von Bedeutungen und komplexen Bedeutungszusammenhängen selbst bewältigt, indem es anhand genuin konnektionistischer Eigenschaften wie Selbstorganisation und Constraint Satisfaction die erforderlichen Leistungen erbringt.

Im Unterschied zu den vielen anderen Repräsentationsformaten, die wir zum Satzverstehen vorgestellt haben (vgl. Kap. 4), bringt dieses reduzierte Format natürlich eine starke Beschränkung der Möglichkeiten mit sich. So ist die Komplexität der referenziellen Situationen stark begrenzt, was sich direkt auf den Skopus möglicher Fragestellungen und Gegenstände auswirkt, die mit diesem Modell sinnvoll simuliert werden können. Viele semantische Gegenstandsbereiche (etwa Handlungsarten, thematische Rollen, gegenseitige semantische Beschränkungen und stereotypische Informationen) bleiben ausgeklammert. Allerdings liegt es nicht im Interesse dieses Forschungsprojekts, großskalierte Anwendun-

---

und die Verarbeitung ihrer räumlichen Orientierung wurde berücksichtigt (zu frühen konnektionistischen Untersuchungen hierzu vgl. Rueckl, Cave & Kosslyn, 1989; Jacobs, Jordan & Barto, 1990).

gen mit einem umfassenden Skopus zu entwickeln. Vielmehr sollen generelle Mechanismen der Abbildung zwischen Form und Bedeutung und der Indizierung perzeptiver Derivate durch sprachliche Sequenzen modelliert werden. Und genau diesem Ziel ist auch die Reduziertheit der von uns gewählten Repräsentation und der dafür implementierten Architektur geschuldet – im Sinne eines fundamentalistischen Modells (Kello & Plaut, 2003; Plaut, 2002).

### 10.1.2.2. Sprachliche Stimuli

Neben den visuellen Stimuli müssen wir die sprachlichen Stimuli spezifizieren. Wir beschreiben an dieser Stelle zunächst das Material für die Simulation SynSem I und verschieben die Diskussion der wesentlich detaillierteren sprachlichen Stimuli für SynSem II auf Kapitel 10.3.1. Die Sätze – sofern man sie als solche bezeichnen möchte – umfassen jeweils nur drei bis vier Wörter sowie ein Startsymbol (beginning of sentence, BOS) und ein Endsymbol (end of sentence, EOS) im Sinne von Interpunktionszeichen. Neben den Symbolen BOS und EOS enthält das Lexikon acht Wörter für die unterscheidbaren räumlichen Relationen, vier Wörter für die unterscheidbaren Objekte, ein Verb (*ist*) und ein Demonstrativum (*das*). Die Mikrogrammatik ermöglicht die folgenden Satzstrukturen: Einfacher Deklarativsatz (Bsp. 33a), normaler Hauptsatz (Bsp. 33b) und Hauptsatz mit topikalisiertem Adverb (Bsp. 33d). Da die mit den Sätzen 33b und 33d beschriebenen Relationen auch genau umgekehrt beschrieben werden können (wenn A links von B ist, dann ist auch B rechts von A), ist für diese beiden Satzstrukturen die jeweilige Inversion möglich (Bsp. 33c und 33e).

- (33) a. BOS das ist ObjektA EOS (*Satztyp: DO*)  
 b. BOS ObjektA ist rechts\_unter ObjektB EOS (*Satztyp: ORO*)  
 c. BOS ObjektB ist links\_über ObjektA EOS (*Satztyp: ORO*)  
 d. BOS links\_über ObjektA ist ObjektB EOS (*Satztyp: ROO*)  
 e. BOS rechts\_unter ObjektB ist ObjektA EOS (*Satztyp: ROO*)

Die Sätze beschreiben die parallel dargebotene visuelle Situation und beziehen sich dabei auf den Aspekt der Situation, der mit der Target-Repräsentation erfasst wird. Dabei liegt die jeweilige statische Situation immer so lange an, bis der korrespondierende sequenziell dargebotene Satz abgeschlossen ist. Die visuelle Situation liegt also immer über mehrere Zeitschritte an. Jeder Situationsaspekt kann potenziell mit allen der fünf möglichen Satzstrukturen beschrieben werden. Setzen wir etwa in die Sätze unter Beispiel 33 für ObjektA die PIPE und für ObjektB den SLASH ein, erhalten wir fünf verschiedene Möglichkeiten, um mit den dabei entstehenden Sätzen auf den gewählten Situationsaspekt in Abbildung 10.2 zu referieren. Dadurch, dass der Satz immer auf den als Target gewählten Situationsausschnitt referiert, soll das Netzwerk lernen, eine sprachliche Auswahl zu treffen. Die Gestaltung der visuellen und sprachlichen Stimuli und die hierfür zu etablierende Abbildung einer Eins-zu-eins-Korrespondenz erfüllt die Voraussetzungen dafür, dass das Netzwerk das Problem der referenziellen Indeterminiertheit bei potenziellen Viele-zu-viele-Korrespondenzen lösen muss: Auf sprachlicher Seite besteht die mehrfache Realisierungsmöglichkeit beschreibender Sätze sowie die Tatsache, dass es sich dabei um

Mehrwortäußerungen handelt. Auf visueller Seite bestehen immer mehrere Möglichkeiten, welches der dargebotenen Objekte und welche der dazwischen bestehenden Relationen sich auf welches Wort beziehen.

### 10.1.3. Aufgaben-Spektrum und Integration

Kehren wir zurück zu den erforderlichen Leistungen des Netzwerks. Wir haben in den vorausgehenden Unterkapiteln die Architektur und die beiden Routen der visuellen und der sprachlichen Verarbeitung besprochen. In der Besprechung der Stimuli haben wir die in den beiden Routen erforderlichen Transformationsleistungen weitergehend erläutert und erste Aspekte der Integration der beiden Routen aufgeführt. Dies führen wir im folgenden Unterkapitel weiter. Wir hatten festgestellt, dass die Informationen gleichzeitig verarbeitet werden: Der visuelle Input wird dabei so lange eingespeist, bis der komplette korrespondierende Satz dargeboten wurde. Wir hatten auch festgestellt, dass sich die beiden Routen einen gemeinsamen Hidden-Layer – den Integrations-Layer – teilen und über diesen ihre jeweiligen Transformations-Aufgaben leisten müssen. Auf die getrennt betrachteten Routen bezogen, leistet der Integrations-Layer zum einen die Erfassung sequenzieller Abhängigkeiten und Konkurrenz für die Verarbeitung der sprachlichen Route. Zum anderen leistet er die kategoriale und prototypische Erfassung der Merkmale der jeweiligen visuellen Situation. Diese Leistungen müssen auch über den gemeinsamen Layer hinweg erbracht werden. Hierzu müssen in irgendeiner Form gemeinsame interne Repräsentationen auf diesem Layer entwickelt werden. Unserer Ansicht nach sind diese gemeinsamen Repräsentationen sowie die Gewichte, die die spezifischen Aktivierungsmuster des Hidden-Layers erzeugen, das Kernstück des Verstehens bzw. der Bedeutungskonstruktion. Über diese Repräsentationen muss die Indizierung der nicht-sprachlichen Repräsentationen durch die Repräsentationen sprachlicher Form erfolgen.

In der ersten Beschreibung der Architektur wurde möglicherweise suggeriert, dass die Propagierung der Aktivierung und die rückwärtsgerichtete Propagierung des Fehlers jeweils getrennt für die Routen erfolgt und unterscheidbar bleibt. Dies ist natürlich nicht der Fall. Zum einen interagieren die beiden Routen ab dem gemeinsamen Integrations-Layer miteinander. Dabei werden die Aktivierungen der visuellen Route genauso durch die Route zur Wortvorhersage weitergeleitet und auf die dort geleistete Aufgabe der Wortvorhersage bezogen wie die Aktivierungen der sprachlichen Route. Und die Aktivierungen der sequenziell dargebotenen sprachlichen Informationen sowie ihrer durch den rekurrenten Layer kontextualisierten Repräsentationen werden genauso auf die Target-Aktivierungen der visuellen Route abgebildet wie der statische visuelle Input. Damit wirken beide Routen gleichberechtigt auch auf die Transformations-Aufgaben der jeweils anderen Route ein. Durch die Rückwärtspropagierung des Fehlers und die gemeinsame Anpassung der Gewichte für das gesamte Netzwerk<sup>3</sup> wird der Einfluss der jeweiligen Informationen als Constraint für die Generierung der erforderlichen Target-Aktivierung in die Gewichte eingeschrie-

---

<sup>3</sup>Wie wir in Kapitel 2.2.1.1 erläutert haben, werden beim Backpropagation-Algorithmus die Fehler zurückpropagiert, um auf dieser Grundlage die Gewichte in mehrschichtigen Netzwerken anzupassen. Bei Backpropagation Through Time werden zudem die den internen Layern zugewiesenen Fehler über mehrere Zeitschritte aufsummiert. Die Anpassung der Gewichte erfolgt kumulativ für die Fehler dieser Zeitschritte zu einem gemeinsamen Zeitpunkt.

ben. Der zurückpropagierte Fehler fällt für die beiden Routen auf dem Integrations-Layer zusammen und wird von dort bis zu den Input-Layern weiterpropagiert. In diesem gemeinsamen Layer müssen die Routen Repräsentationen aushandeln, die die Erfüllung beider Transformations-Aufgaben gewährleistet.

Auch wenn die Routen als getrennte Pfade dargestellt werden, so handelt es sich doch um nur ein Netzwerk, dessen Input- und Outputmuster eben komplexer gestaltet sind und über einen dynamischen bzw. veränderlichen und einen statischen Anteil verfügen. Revidieren wir aus dieser Perspektive die oben dargestellten Aufgaben, die nun weniger anhand der Routen und mehr anhand der zu leistenden Transformationen aus dem gemeinsamen Input bestimmt werden müssen:

- Das Netzwerk muss zwei synchrone Input-Quellen verarbeiten. Über die visuelle Route werden Informationen aus einem zweidimensionalen Unit-Gitter eingespeist, über die sprachliche Route werden sequenziell lokalistisch repräsentierte Wörter dargeboten.
- Um die visuellen Informationen nutzen zu können, müssen (a) die mitunter orthogonalen Punktmuster<sup>4</sup> als einheitliche Objekte identifiziert werden, (b) die Objekte voneinander unterschieden werden, (c) Aktivierungsverteilungen als räumliche Verortungen interpretiert werden, (d) die Objekte den Positionen zugewiesen werden und (e) die Verortungen als relational zueinander erkannt werden.
- Um die sprachlichen Informationen nutzen zu können, müssen (a) Wahrscheinlichkeiten von Wortabfolgen bzw. Übergangs-Wahrscheinlichkeiten abgeleitet werden, (b) Kollokationen von Wörtern erkannt werden, (c) anhand dieser Informationen Wortklassen gebildet werden und (d) eventuell Generalisierungen über Abfolge-Regelmäßigkeiten abstrakter Wortklassen geleistet werden.<sup>5</sup>
- Die Transformations-Aufgaben erfordert für den sprachlichen Output die Abbildung des Inputs auf eine Sequenz von Wörtern und für den visuellen Output die Abbildung des Inputs auf ein statisches Muster. Sowohl die jeweils möglichen Wörter im Verlauf der Sequenz als auch die im statischen visuellen Output abzubildenden Objekte und Relationen werden durch Informationen aus beiden Informations-Routen bedingt.
- Damit die Informationen wechselseitig sinnvoll genutzt werden können, müssen die einzelnen Wörter, die Zusammenhänge verschiedener Wörter in ihrer Abfolge und die strukturellen Zusammenhänge zwischen den Wörtern mit den identifizierten Eigenschaften des visuellen Inputs integriert werden: (a) So müssen zum Beispiel die Nomen in Beziehung zu den Objekten gebracht werden, (b) des Weiteren spielt es eine entscheidende Rolle, ob das Nomen vor oder nach dem Adverb folgt, da sich nur aus dieser Reihenfolge ergibt, ob das benannte Objekt etwa links oder rechts von einem anderen Objekt steht und (c) schließlich muss aus den Informationen

<sup>4</sup>Die Objekte sind im Input-Gitter frei beweglich. Übertragen auf das binäre Vektor-Format bedeutet dies aber, dass zum Beispiel 00001100 dasselbe bedeuten kann wie 01100000. Ein und dasselbe Objekt kann also durch gänzlich orthogonale, d.h. maximal unähnliche Vektoren dargestellt werden. Zugleich können sich zwei verschiedene Objekte besetzte Vektor-Stellen teilen, wenn sie auf der gleichen Position im Gitter erscheinen, und sind damit ähnlicher zueinander als identische Objekte auf verschiedenen Positionen.

<sup>5</sup>Allerdings ist es eher unwahrscheinlich, dass bei einer solch primitiven Grammatik und einem so kleinen Lexikon abstrakte syntaktische Generalisierungen geleistet werden müssen.

des sequenziellen Kontexts bestimmbar sein, welche Bestandteile bereits verarbeitet wurden, um bereits erfasste und noch zu erfassende Aspekte der Situation zu unterscheiden und um korrekte grammatikalische Abfolgen sicherzustellen.

- Die visuellen Informationen dienen bei der Wortvorhersage-Aufgabe als semantischer Constraint für die in der gegebenen Situation möglichen Folgewörter und unterstützen möglicherweise den Erwerb syntaktischer Klassen und Regelmäßigkeiten im Sinne des semantischen Bootstrapping.
- Die sprachlichen Informationen dienen bei der visuellen Muster-Assoziation als linguistischer Constraint für die Auswahl der Objekte bzw. der relevanten Relation und unterstützen möglicherweise durch sequenzielle Kookkurrenz-Informationen den Bedeutungserwerb.
- Die wechselseitigen Constraints der verschiedenen Informationsstrukturen und ihr Anteil an der zu erfüllenden Aufgabe bilden sich in den Gewichtsvektoren zwischen den Layern ab. Wir nehmen an, dass sich diese Prägung der internen Strukturen in der Herausbildung gemeinsamer interner Repräsentationen verdichten wird. Diese erwarten wir vor allem im rekurrenten Integrations-Layer als zentralem Ort der (sequenziellen) Kontextualisierung, der Entwicklung von zentral transformationsrelevanten Repräsentationen und der Aushandlung zwischen den Constraints der beiden Routen.

### 10.1.3.1. Drei Test-Modi

Wir haben das erforderliche Leistungsspektrum und die dabei erwarteten Interaktionen skizziert. Wie sieht aber die tatsächliche Performanz der Netzwerke aus, welchen Anteil an der Transformations-Leistung haben die beiden Informationsquellen tatsächlich, und wie beeinflussen sie sich gegenseitig? Die gewählte Architektur legt drei Test-Modi nahe, um die Performanz der Netzwerke eingehender zu untersuchen: (1) Test mit Input für beide Routen, (2) Test mit Input ausschließlich für die sprachliche Route und (3) Test mit Input ausschließlich für die visuelle Route. In allen drei Fällen kann die Output-Aktivierung beider Routen untersucht werden: Wie präzise ist die Wortvorhersage, wenn nur der sprachliche Input gegeben ist, und wie gestaltet sich der visuelle Output? Bildet sich ein sequenziell sprachlicher Output, wenn nur visueller Input gegeben ist, und welche Merkmale zeigen sich dann für den visuellen Output? Welchen Einfluss haben bei beidseitig gegebenem Input die visuellen Informationen auf die sprachliche Wortvorhersageleistung und umgekehrt die sprachlichen Informationen auf die visuelle Muster-Assoziation sowie die hierbei erforderliche Prototypisierung der Situation? Die drei möglichen Test-Modi können in Korrespondenz zur menschlichen Sprachverarbeitung betrachtet werden:

**Situierte Sprachverarbeitung** Mit den Informationen beider Input-Routen simulieren wir die situationsbezogene Verarbeitung und Interpretation des Satzes. Diese Ausgangssituation könnte man auch mit modalitätsübergreifendem Priming und Lenkung visueller Aufmerksamkeit (und entsprechenden empirischen Daten) in Verbindung bringen.

**Sprachverstehen** Mit ausschließlich sprachlichem Input simulieren wir entkoppeltes bzw. situationsunabhängiges Sprachverstehen, das Verstehen einer raum-zeitlich abwe-

senden Situation. Aus der Perspektive der Embodied Cognition entspricht dies dem Verstehensprozess als sprachlich induzierten Enactment modaler Erfahrungsderivate. Diese Ausgangssituation sollte sowohl auf die Wortvorhersage als auch auf die Konstruktion der visuellen Target-Repräsentation Auswirkungen haben.

**Sprachproduktion** Mit ausschließlich visuellem Input simulieren wir die Produktion einer Situationsbeschreibung. Für die Simulation von Sprachproduktion müssen wir eine Erweiterung der Architektur einführen, die im folgenden Abschnitt erläutert wird. Auch hier sollte sich die freiere Ausgangssituation sowohl auf die Wortvorhersage als auch auf die visuelle Muster-Assoziation auswirken.

Für die Aufgabe der Sprachproduktion (als Generierung situationsbezogener Wort-Sequenzen) wird die reine Darbietung des visuellen Inputs nicht genügen, denn er bietet keine Information zur sequenziellen Orientierung, d.h. keine Rückmeldung dazu, an welcher Stelle in der Sequenz sich die entstehende Situationsbeschreibung gerade befindet. Um diese sequenzielle Rückmeldung zu leisten, muss das Netzwerk erweitert werden. Hier kommen wir zum letzten Detail der Architektur: Die Fähigkeit zur selbstüberwachten Produktion von Wort-Sequenzen wurde durch eine einfache Funktion ermöglicht, die eine Feedback-Schleife vom sprachlichen Output-Layer zum sprachlichen Input-Layer herstellt (vgl. Abb. 10.1, zurückführender gestrichelter Pfeil in der linken Hälfte). Dazu wird zuerst die Aktivierungsfunktion des Output-Layers zu einer Winner-takes-all-Funktion modifiziert. Diese ermittelt aus der Aktivierungsverteilung die am stärksten aktivierte Unit, weist dieser Unit den Aktivierungswert 1.0 zu und stellt alle anderen Units des Layers auf eine Aktivierung von 0.0. Dieser lokalistische Vektor, der das vom Netzwerk ermittelte wahrscheinlichste Folgewort bestimmt, wird über eine Eins-zu-eins-Verknüpfung (mit auf den Faktor 1.0 fixierten Gewichten) auf den sprachlichen Input-Layer kopiert. Hier dient der Vektor im nächsten Zeitschritt als neuer Input. Das Netzwerk erhält also zu Beginn keinen externen Input, vielmehr entwickelt es nach der ersten Satzposition durch den Feedback aus dem Output seinen eigenen Input im Sinne eines internen Verbalisierens. Die aus dem internen Input abgeleitete Vorhersage des nächsten Worts ergibt jeweils den nächsten Input. Auf diese Weise muss das Netzwerk aus dem eigenen internen Wissen über wohlgeformte Sequenzen eine kohärente Wort-Sequenz bilden. Ein solches Vorgehen zur Simulation von Sprachproduktion als selbstinduzierte Generierung von Wortketten wurde bereits in anderen Simulationen in einem gewissen Rahmen erfolgreich angewendet (Dell, Chang & Griffin, 1999; Christiansen & Chater, 1999b; Rohde, 2002).

Ähnlich wie im CSCP-Modell (Rohde, 2002) und im Unterschied zu den anderen beiden genannten Modellen<sup>6</sup> erfolgt die Sprachproduktion aber nicht völlig ungesteuert. Das Netzwerk soll nicht willkürliche Wort-Sequenzen bilden, die nur grammatische Wohlgeformtheit erfüllen. Vielmehr sollen die generierten Sätze einen bestimmten semantisch spezifizierten Inhalt transportieren. Die Kontrolle der Sprachproduktion erfolgt über den visuellen Input, der eine zu interpretierende bzw. sprachlich zu beschreibende Situation vorgibt. Das Netzwerk erhält also einerseits nach dem initialen sprachlichen Startsignal keinen weiteren externen sprachlichen Input. Andererseits liegt aber durchgehend eine

---

<sup>6</sup>Im Modell von Christiansen und Chater (1999b) etwa wurden durch das Netzwerk beliebige Wort-Sequenzen generiert. Satzlänge, Satzstruktur und grammatische Markierungen waren willkürlich, vor allem aber wurden keine semantischen Beschränkungen oder gar wiederzugebende semantische Inhalte auferlegt. Einziges Kriterium blieb die Grammatikalität der Sequenzen.

spezifische visuelle Situation an. Wir erwarten, dass der generierte Satz Aspekte der in dieser Situation dargebotenen Sachverhalte wiedergibt. Damit müssen in der Simulation der Sprachproduktion nicht nur Kriterien der grammatischen, sondern auch der semantischen Kohärenz erfüllt werden.

## 10.2. Modellierung SynSem I

Das folgende Kapitel stellt die Simulationen mit der Architektur und den Stimuli des Projekts SynSem I dar. Das Modell soll lernen, aus der parallel-interaktiven Verarbeitung der korrespondierenden Stimuli und der synchron zu erfüllenden Aufgaben Leistungen zu erzielen, die die bloße Bewältigung der Einzelaufgaben weit überschreitet. Wir demonstrieren, dass das Netzwerk lernt, sinnvolle interne Repräsentationen zu entwickeln, und die gegenseitig beschränkenden Informationen der interagierten Stimuli und Aufgaben zu nutzen. Dabei simulieren wir zum einen den Vokabelerwerb bei potenzieller Viele-zu-viele-Korrespondenz sowie den Satzverarbeitungserwerb als Kombination lexikalischer Antizipation und inkrementeller Konstruktion korrespondierender nicht-sprachlicher Situationsmodelle. Wir untersuchen kontrastierend die Performanz in den verschiedenen Test-Modi. Dabei können wir demonstrieren, wie sich im Modell durch die Informationsintegration und Aufgabeninteraktion weitergehende nicht explizit implementierte Eigenschaften entwickeln. Diese können im Sinne weitergehender Leistungen wie situationaler Inferenz, Lenkung visueller Aufmerksamkeit oder modalitätsübergreifendem Priming bei der lexikalischen Antizipation interpretiert werden. Diese Eigenschaften entwickeln sich emergent als Nebenprodukt multipler Constraint-Satisfaction-Prozesse und nicht wie in vielen früheren Modellen durch eine explizite Implementierung entsprechender Mechanismen.

Wir legen zunächst einige Details zu den verwendeten Stimuli dar und erläutern den Trainings-Aufbau und die verwendeten Parameter für die Simulationen. Anschließend benennen wir die erhobenen und verglichenen Faktoren und Bedingungen und erläutern die Aufbereitung der Performanz-Maße und der Kennwerte. Nach den jeweiligen allgemeinen Ergebnissen verdeutlichen wir anhand von Diagrammen und Beispielen die spezifischen Modell-Eigenschaften.

### 10.2.1. Stimuli und Simulationen

#### 10.2.1.1. Trainings- und Test-Set

Den Stimuli des visuellen Inputs liegt ein zweidimensionales retinales Unit-Gitter von 9x9 Units (insg. 81 Units) zugrunde.<sup>7</sup> Wir verwenden die vorgestellten vier unterschiedlichen Objekte. Von diesen können je visueller Situation eines, zwei oder drei gleichzeitig auftreten, wobei sich gleichzeitig auftretende Objekte unterscheiden müssen. Die Objekte dürfen sich frei auf dem retinalen Gitter verteilen und sich dabei auch berühren, aber

---

<sup>7</sup>Sowohl für den Input- als auch den Output-Layer der visuellen Route ermöglicht das verwendete Tool auch andere Größen. Für die prinzipielle Untersuchung der Netzwerkleistung genügt aber diese relativ kleine Größe der Layer, zumal diese hinsichtlich der Zeit-Ersparnis bei der Simulation pragmatischer ist.

nicht überlappen. Bei vollständiger Kombination und Verteilung aller möglichen Objekt-Konstellationen ergeben sich 130.708 kombinatorisch mögliche Input-Situationen. Aus diesem Pool möglicher Situationen haben wir eine Untermenge von knapp 54% (insgesamt 70.000 Situationen) ausgewählt und mit einem korrespondierenden Satz sowie einem korrespondierenden Target-Vektor versehen.

Den idealisierten und prototypisierten Target-Vektoren liegt ein Output-Gitter von  $7 \times 7$  Units (insg. 49 Units) zugrunde. Auf dieses Output-Gitter werden die Input-Situationen abgebildet. Wenn nur ein einzelnes Objekt im Input gegeben ist, wird dieses Objekt genau in der Mitte des Gitters abgebildet. Der korrespondierende Satz beschreibt die Input-Situation deklarativ mit einem Demonstrativpronomen (Satztyp DO, vgl. Bsp. 33).<sup>8</sup> Wenn im Input zwei Objekte gegeben sind, werden diese in idealisierter Position im Output abgebildet. Dabei werden die Objekte nicht mehr frei verteilt auf dem Output-Gitter platziert, sondern auf festen vorgegebenen Positionen an den entgegengesetzten Enden des Gitters, mit denen die räumliche Relation klar wiedergeben wird (vgl. Abb 10.2). Der korrespondierende Satz beschreibt die in der Input-Situation gegebene Relation der Objekte (Satztypen ORO und ROO). Bei drei Objekten im Input werden zwei der gegebenen Objekte ausgewählt und im Output idealisiert positioniert. Hier beschreibt der korrespondierende Satz die Relation der ausgewählten Objekte.

Für das Test-Set wurde aus den nicht im Training erscheinenden 46% der im Input möglichen Objekt-Konstellationen eine Untermenge von 600 Situationen ausgewählt:  $300 \times 3$  Objekte,  $200 \times 2$  Objekte,  $100 \times 1$  Objekt.<sup>9</sup> Damit wurde sichergestellt, dass zumindest das Input-Material in diesen Konstellationen nicht bereits im Training zugegen war, und das Netzwerk zur Verarbeitung dieser Stimuli tatsächlich eine Generalisierungsleistung erbringen muss. Dieses Material wird in den oben besprochenen drei verschiedenen Test-Modi dargeboten: (1) Sowohl für die visuelle als auch die sprachliche Route wird Input dargeboten. (2) Nur für die sprachliche Route wird Input dargeboten, auf dem visuellen Input-Layer erhalten alle Units eine Aktivierung von 0.0. (3) Nur für die visuelle Route wird Input dargeboten, auf dem sprachlichen Input-Layer erhalten alle Units eine Aktivierung von 0.0. Im dritten Test-Modus erhält der sprachliche Output-Layer die Aktivierungsfunktion Winner-takes-all, und das Netzwerk wird um die besprochene Feedback-Schleife erweitert.

### 10.2.1.2. Simulationen und Parameter

Die Netzwerke wurden für 20 Epochen trainiert. Eine Epoche umfasst das gesamte Set von 70.000 Situations-Satz-Paaren. Dabei wurde auch im Training die Unterscheidung der Arbeits-Modi umgesetzt: Die Hälfte der Paare wurde mit vollständigem Input dargeboten, jeweils ein Viertel der Paare entweder nur mit sprachlichem oder nur mit visuellem Input. Diese Vorgehensweise ist nötig, um zu vermeiden, dass mit der Aktivierung der jeweiligen

<sup>8</sup>Die Situationen mit nur einem Objekt erscheinen im Trainings-Set mehrfach, da sie im Vergleich zu den anderen Situationen sehr selten sind. Da es im Vergleich zu den Situationen mit mehreren Objekten kombinatorisch nur sehr wenige Situationen gibt, wurden die Situationen mit nur einem Objekt zudem vollständig ins Trainings-Set übernommen. Wir haben diese Situationen mit zehnfacher Häufigkeit dargeboten (insg. 1960 Trainings-Beispiele).

<sup>9</sup>Ausschließlich bei den Situationen mit nur einem Objekt haben wir Situationen im Test-Set verwendet, die bereits im Trainings-Set erschienen.

Units von 0.0 assoziiert wird, dass das jeweilige Wort oder die jeweilige visuelle Position nicht besetzt ist. Denn dies könnte sich in den reduzierten Test-Modi so auswirken, dass die fehlende Input-Aktivierung als diskonformierender Stimulus für die jeweilige Target-Repräsentation interpretiert und ein Überkreuz-Test der Routen in der Folge unmöglich wird.

Wie oben beschrieben, haben wir verschiedene Versionen der Architektur trainiert, die sich in der Anzahl zusätzlicher Kompressions-Layer unterscheiden. Dabei zeigten sich nur minimale quantitative und praktisch keine qualitativen Unterschiede. Wir beschränken uns im Folgenden auf die Besprechung der vollständigen Architektur mit zwei Kompressions-Layern für die visuelle Route, mit der die besten Ergebnisse erzielt wurden. Auch die Größe der Hidden-Layer hatte innerhalb sinnvoller Unit-Mengen keinen Einfluss auf die Performanz. Die Konstellation der besprochenen Netzwerke umfasst Layer folgender Größe: sprachlicher Input und Output jeweils 16 Units, visueller Input 81 Units, visueller Output 49 Units, Integrations-Layer (sowie Kontext-Layer) 40 Units, die beiden Kompressions-Layer ebenfalls jeweils 40 Units. Der Output-Layer der sprachlichen Route erhält eine Softmax-Funktion, die die Summe des Outputs auf 1.0 normalisiert.<sup>10</sup> Für die Umsetzung der Simulationen haben wir wieder die Modellierungs-Software LENS (Rohde, 1999a) verwendet. Die Netzwerke wurden mit Backpropagation Through Time (Werbos, 1990) mit einer Batch-Größe von 1 trainiert (d.h. nach jedem vollständigen Satz werden die Gewichte upgedatet), der Fehler wurde dabei anhand Kreuzentropy bemessen. Nach jedem Situations-Satz-Paar wurde der Kontext-Layer gelöscht.

Wie bei den Architektur-Spezifikationen haben wir auch bei den Trainings-Parametern ein Spektrum sinnvoller Werte getestet: Für die Lernrate wurden Werte zwischen 0.05 und 0.2 verwendet, für das Momentum Werte zwischen 0.0 und 0.6, für das Spektrum möglicher initialer Gewichts-Werte Obergrenzen zwischen 0.1 und 1.0. Auch gegenüber den verschiedenen Trainings-Parametern zeigte sich das Modell sehr robust. Es gab lediglich minimal bessere Werte für höhere Parameter-Einstellungen. Die Ergebnisse, die wir im Weiteren darstellen, beruhen auf diesen Einstellungen: Für die Lernrate wurde  $\epsilon = 0.2$  gewählt, für das Momentum der Faktor  $momentum = 0.6$ , die initiale Randomisierung der Gewichte lag innerhalb der Obergrenze von  $initialweightrange = 1.0$ . Wir greifen in der Besprechung auf die Ergebnisse einzelner Netzwerke zurück. Die Validität der Ergebnisse ist über die qualitativ identische Performanz in anderen Trainings-Konstellationen gewährleistet und wird durch die Robustheit des Modells gegen eine Vielzahl an Parametern und Architektur-Spezifikationen bestätigt.

### 10.2.2. Auswertungsverfahren

Bevor wir die Ergebnisse präsentieren, erläutern wir an dieser Stelle die aufgenommenen Faktoren und die berechneten Kennwerte aus den Output-Dateien. Wir beginnen mit den Faktoren und deren Bedingungen, anschließend benennen wir die unterschiedlichen Kennwerte für Performanzaspekte der Wortvorhersage und der Musterassoziation und erläutern die Auswertungsverfahren, über die die Kennwerte für die gespeicherten Aktivierungs-

---

<sup>10</sup>Eine entsprechende Anpassung des visuellen Output-Layers wäre hingegen nicht sinnvoll, da es hier keine vorab bekannte Ober- oder Untergrenze der Summe der Aktivierungen gibt. Zudem wollen wir die hier gegebenen Aktivierungen nicht im Sinne der Verteilung absoluter Wahrscheinlichkeiten interpretieren.

muster aus den Testdurchläufen (d.h. die Output-Dateien für unsere Test-Sets) berechnet wurden. Für die Analysen wurden die folgenden Faktoren unterschieden:

- Netzwerk-Nummer (1-5)
- Trainingsparameter (Lernrate, Momentum, initiale Gewichtsverteilung)
- Epoche (0-20)
- Beispiel-Nummer (fortlaufende Nummerierung der Sätze)
- Wort-Nummer (fortlaufende Nummerierung der Wörter innerhalb eines Satzes)
- Testmodus (Verstehen, Produktion, situiertes Verstehen)
- Aktuelles Wort (vgl. Lexikon der Grammatik)
- Satztyp des aktuellen Satzes (vgl. Satztypen der Grammatik)
- Anzahl aktuell gegebener Input-Objekte (1-3)

Folgende Liste gibt die Kennwerte für die Komponenten Wortvorhersage und Musterassoziation wieder. Anhand dieser Werte werden wir in den Ergebnis-Abschnitten vergleichende Diagramme für Mittelwerte für die verschiedenen Bedingungen erstellen.<sup>11</sup> Die ersten drei Werte dienen als Maß zur Beurteilung der Wortvorhersage-Performanz, die folgenden fünf Werte beurteilen verschiedene Aspekte der Muster-Assoziation (Korrektheit der Objekte, der Relation sowie der gesamten Szene). Zu den jeweiligen Werten wird erläutert, über welche Berechnungsschritte und auf welcher Basis sie erhoben wurden.<sup>12</sup> Wir haben dabei alle Werte so ermittelt oder umgeformt, dass sie als Fehlerwert interpretiert werden können (d.h. je geringer desto besser).

**1-CatCor Syn** Für dieses Fehlermaß wird die am stärksten aktivierte Unit der Wortvorhersage ermittelt und überprüft, ob das entsprechende Wort zur selben Wortklasse gehört wie das Ziel-Wort. Wir haben für die beiden Simulations-Reihen Nomen, Verben, Adverbien (für die Relationen), Demonstrativa, Satzmarker und Interrogativa unterschieden. Bei korrekter Wortklasse wird der Wert 1.0 vergeben, bei falscher Wortklasse der Wert 0.0; mit der Differenz des Werts zu 1.0 erhalten wir den entgegengesetzten Fehlerwert.

**EukDist Syn** Wir berechnen die euklidische Distanz zwischen dem Output-Vektor und dem Target-Vektor der Wortvorhersage (vgl. Kap. 3.2.1.1). Ein Wert von 0.0 entspricht einer minimalen Distanz, nach oben ist der Wert nicht begrenzt.<sup>13</sup>

<sup>11</sup>Unter den erhobenen Fehlermaßen haben sich nicht alle als sinnvoll oder aussagekräftig erwiesen. So wurden z.B. nach dem Vorbild des GPE (vgl. MacDonald & Christiansen, 2002) die Kennwerte VglQuot Syn und VglQuot Sem entwickelt und erhoben. Da diese Kennwerte aber keinerlei Zusatzinformationen erbrachten und kaum einen Vergleich zwischen verschiedenen Ergebnissen ermöglichten, haben wir auf diese verzichtet. Im Falle der Sprachkomponente beruht der erhobene Wert zudem auf einem Vergleich mit einer willkürlich gewählten Toleranzschwelle, die den Wert maßgeblich beeinflusst und damit uninterpretierbar macht.

<sup>12</sup>Die Funktionen zur Auswertung wurden zu großen Teilen von Lars Konieczny und Daniel Müller sowie vom Autor entwickelt und erweitert.

<sup>13</sup>Für die euklidische Distanz und den Fehlerwert 1-Cos ist folgendes anzumerken: Für den Target-Vektor wird immer nur eine Target-Unit aktiviert, während der Output-Vektor alle auf dieser Satzposition möglichen Units aktiviert. Folglich haben größere Wortklassen immer einen höheren Fehler als kleine Wortklassen, da hinsichtlich der Kollokation korrekt aktivierte Units als Abweichung vom Target-Vektor

**1-Cos Syn** Für diesen Fehler wird die Ähnlichkeit der Output- und Target-Vektoren mit dem Cosinus verglichen. Der Betrag wird von 1.0 subtrahiert, um einen Fehlerwert zwischen 0.0 (maximal ähnlich) und 1.0 (maximal unähnlich) zu erhalten.

**LocScore Sem** Für diesen Fehlerwert wird der Output-Layer der visuellen Route sozusagen abgescannt: In einem Fenster von jeweils  $3 \times 3$  Units (entsprechend der räumlichen Ausdehnung der möglichen Objekte) werden die summierten Aktivierungen aller möglichen Objekt-Positionen auf dem zweidimensionalen Unit-Gitter verglichen. Die beiden aktivsten Regionen bestimmen die vom Netzwerk zugewiesene räumliche Relation der Objekte in ihrer horizontalen und vertikalen Position. Anhand des Cosinus wird die Ähnlichkeit der Regionen zu den erwarteten Regionen des Target-Vektors bemessen, das Ergebnis wird von 1.0 subtrahiert. Damit erhält der LocScore Sem einen Wert zwischen 0.0 (Übereinstimmung der Relation) und 1.0 (keine Übereinstimmungen).

**1-ObjCor Sem** Die Aktivierungsverteilung der ermittelten Objekt-Positionen wird anhand des Distanzmaßes der euklidischen Distanz mit dem Aktivierungsmuster der möglichen Objekte verglichen. Die jeweils ähnlichsten Objekte werden als vom Netzwerk ermittelte Objekte erfasst. Durch den Vergleich mit den Ziel-Objekten des Target-Vektors wird die Korrektheit der ermittelten Objekte bestimmt: Jedes korrekte Objekt erhält einen Wert von 0.5. Die Summe der Werte ergibt einen Gesamtwert von 0.0 bei keiner Übereinstimmung, von 0.5 bei einem korrekten Objekt und von 1.0, wenn beide Objekte korrekt sind. Durch die Subtraktion von 1.0 wird der Wert als Fehlerwert interpretiert.

**PosRel Sem** Die ermittelte Objekt-Relation im Output der visuellen Route (BestRel) wird mit den möglichen Relationen aller Objekte der jeweiligen Input-Situation verglichen.<sup>14</sup> Wenn die als am besten ermittelte Relation mit einer der möglichen Relationen übereinstimmt, wird dem Fehler PosRel der Wert 0.0 zugewiesen, ansonsten der Wert 1.0.

**EukDist Sem** Euklidische Distanz zwischen dem Output-Vektor und dem Target-Vektor der visuellen Route. Anders als bei der Wortvorhersage-Komponente kann hier der Output mit dem gesamten Zielzustand verglichen werden. Allerdings wird nicht berücksichtigt, ob die Output-Aktivierung eine alternative und (zumindest temporär) ebenfalls korrekte Objekt-Relation erfasst, solange die bisher angelegten Wörter des sprachlichen Inputs noch mehrere Interpretationen der Situation zulassen.

**1-Cos Sem** Entsprechendes gilt für den Cosinus dieser Vektoren. Der Betrag wird hier von 1.0 subtrahiert, um einen Fehlerwert zwischen 0.0 und 1.0 zu erhalten.

---

erfasst werden. Dennoch läßt der Fehlerwert einen Vergleich zwischen verschiedenen Bedingungen zu. Zudem sollte durch zusätzliche situationale Inferenzen die Anzahl in Betracht gezogener Wörter einer Wortklasse weiter beschränkt werden – je besser diese Inferenz funktioniert, desto geringer der entsprechende Fehler für diese Fehlermaße.

<sup>14</sup>Bei zwei Objekten gibt es nur eine mögliche Relation, bei drei Objekten ergeben sich drei mögliche Relationen. Diese werden aus einem Protokoll-Dokument entnommen, das bei der Generierung der Trainings- und Test-Files erstellt wird.

### 10.2.3. Ergebnisse

Wir haben die Ergebnisse in verschiedene Unterkapitel unterteilt, in denen die Auswertungen der Performanz für die beiden Leistungsaspekte Musterassoziation und Wortvorhersage nach den unterschiedlichen Test-Modi getrennt dargestellt werden. Unterkapitel 1 enthält die Ergebnisse bei entkoppeltem Sprachverstehen. Unterkapitel 2 enthält die Ergebnisse der Verstehensleistung im Modus des situierten Sprachverstehens. Unterkapitel 3 enthält die Ergebnisse für die Sprachproduktion. Die Analysen erfolgen vorwiegend anhand von Performanz-Beispielen sowie anhand von vergleichenden Diagrammen für die verschiedenen Kennwerte. Über die Performanz-Beispiele für die Aktivierungen im Verlauf einzelner Sätze erläutern wir allgemein beobachtbare Leistungsaspekte. Da wir für die situationale Repräsentation anstatt willkürlicher Punktmuster oder semantischer Feature-Repräsentationen einfache räumliche Situationen mit konkreten visuell unterscheidbaren Objekten und Verortungen verwendet haben, sind die Leistungsaspekte intuitiv nachvollziehbar. Die Interpretation der Performanzbeispiele belegen wir über die weitergehenden Analysen mit den Kennwerten und vergleichenden Diagrammen für deren Mittelwerte (mit Standardfehler). Anhand der Diagramme kann nachvollzogen werden, dass die erläuterten Einzelbeispiele keine Ausnahmen des Netzwerk-Verhaltens sind, sondern die allgemeine Performanz des Netzwerks widerspiegeln. Zudem lassen sich weitergehende Unterschiede zwischen verschiedenen Bedingungen der Performanz feststellen.

#### 10.2.3.1. Ergebnisse 1: Verstehen (entkoppelt)

Wir verwenden für die folgenden Kennwerte und Diagramme die Ergebnisse der 15. Epoche, da sich ab dieser Epoche keine oder nur marginale Veränderungen zeigen. Betrachten wir zunächst die Kennwerte im Test-Modus mit ausschließlich sprachlichem Input. Wir verwenden dafür das arithmetische Mittel und die Standardabweichung der Fehlerwerte über alle Satzpositionen aller Beispiele dieses Modus.

**Globale Kennwerte: Visuelle Route** Folgende Werte wurden ermittelt: Die getrennt ermittelten Werte für Relation und Objekte betragen  $\bar{x}_{\text{LocScore Sem}} = 0.124$  ( $\sigma = 0.176$ ) und  $\bar{x}_{1\text{-ObjCor Sem}} = 0.302$  ( $\sigma = 0.286$ ). Die übergreifenden Fehlerwerte nach Distanz und Ähnlichkeit betragen  $\bar{x}_{\text{EukDist Sem}} = 1.393$  ( $\sigma = 0.871$ ) und  $\bar{x}_{1\text{-Cos Sem}} = 0.31$  ( $\sigma = 0.311$ ). Das Erzielen einer möglichen Relation wird mit  $\bar{x}_{\text{PosRel Sem}} = 0.574$  ( $\sigma = 0.495$ ) bewertet. Diese Kennwerte scheinen zunächst auf eine eher mäßige Leistung hinzuweisen. Die hohe Standardabweichung zeigt zudem die hohe Streuung der Werte. Allerdings ist zu berücksichtigen, dass auf den frühen Satzpositionen nur beschränkte oder überhaupt keine Informationen zur korrespondierenden visuellen Situation gegeben sind. Die Situation muss zuerst nach den Vorgaben der sprachlichen Beschreibung konstruiert werden. Dies wird deutlich, wenn wir zum Vergleich die Kennwerte dieses Modus für die Satzposition heranziehen, nach der alle benötigten Informationen verfügbar waren. Wir verwenden dafür die Mittelwerte über alle Beispiele für den EOS, der das Satzende markiert. Sowohl bei den Relationen als auch bei den Objekten zeigt sich eine Minimierung des Fehlers:  $\bar{x}_{\text{LocScore Sem}} = 0.009$  ( $\sigma = 0.02$ ),  $\bar{x}_{1\text{-ObjCor Sem}} = 0.026$  ( $\sigma = 0.114$ ). Dies spiegelt sich auch in den ähnlichkeits- und distanzbasierten Fehlermaßen und den Kennwerten für die

Gesamtsituation mit  $\bar{x}_{\text{EukDist Sem}} = 0.333$  ( $\sigma = 0.425$ ),  $\bar{x}_{1-\text{Cos Sem}} = 0.0243$  ( $\sigma = 0.053$ ) und  $\bar{x}_{\text{PosRel Sem}} = 0.051$  ( $\sigma = 0.22$ ). Zum Ende des jeweiligen Satzes werden die Situationen weitgehend fehlerfrei abgebildet, falsche Objekt-Relationen kommen beinahe nicht mehr vor.

**Globale Kennwerte: Sprachliche Route** Auf die Performanz der sprachlichen Route wirkt sich das Fehlen des visuellen Inputs anders aus – dies werden wir deutlicher sehen, wenn wir im folgenden Unterkapitel den Unterschied bei gegebenem visuellem Input sehen. Während die Wortklasse dem Kennwert 1-CatCor zufolge weitgehend korrekt vorhergesagt wird, bleiben die anderen Fehlermaße vergleichsweise hoch:  $\bar{x}_{1-\text{CatCor Syn}} = 0.08$  ( $\sigma = 0.271$ ),  $\bar{x}_{\text{EukDist Syn}} = 0.466$  ( $\sigma = 0.484$ ) und  $\bar{x}_{1-\text{Cos Syn}} = 0.32$  ( $\sigma = 0.374$ ). Betrachtet man den globalen Fehler dagegen auf einzelnen Positionen, so fällt auf, dass er für das Verb und den EOS-Marker sehr gering ausfällt. Für den EOS-Marker etwa zeigen sich die folgenden Fehlerwerte:  $\bar{x}_{1-\text{CatCor Syn}} = 0.0$ ,  $\bar{x}_{\text{EukDist Syn}} = 0.004$  ( $\sigma = 0.016$ ) und  $\bar{x}_{1-\text{Cos Syn}} < 0.001$  ( $\sigma = 0.001$ ). Zudem werden auf den anderen Positionen zumeist korrekt entweder Objekte oder Adverbien vorhergesagt – was den niedrigen positionenübergreifenden 1-CatCor-Wert erklärt. Der hohe Fehler der Objekt- und Adverb-Position liegt daran, dass der Output-Vektor eine Wahrscheinlichkeitsverteilung widerspiegelt, der Target-Vektor jedoch nur ein einziges aktuell korrektes Wort erlaubt. Sobald also mehr als ein Wort korrekt ist, müssen ähnlichkeits- und distanzbasierte Fehlermaße ansteigen. Warum dies für SynSem dennoch sinnvolle Kennwerte sind, sehen wir im folgenden Unterkapitel zum situierten Verstehen.

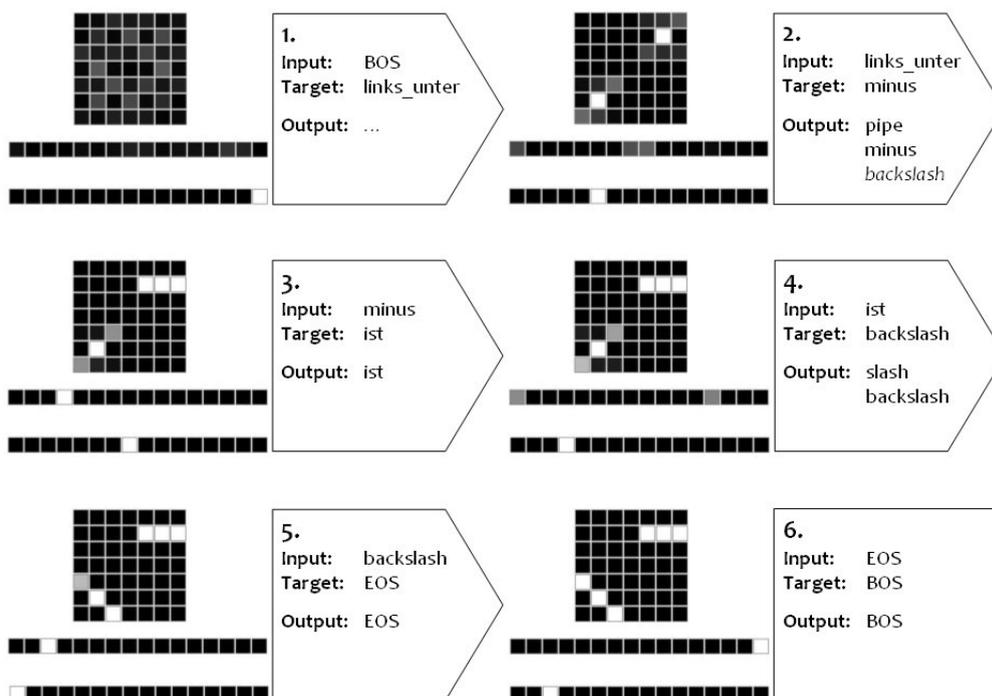


Abbildung 10.3.: SynSem I: Beispiel zum entkoppelten Verstehen eines Satzes und zum inkrementellen Aufbau des Objekt-Relations-Modells.

**Beispiel zum entkoppelten Verstehen** Was bei der Verarbeitung der Sätze beim entkoppelten Verstehen genau geschieht, veranschaulicht das folgende Beispiel in Abbildung 10.3. Die dargestellten Beispiele zeigen aufbereitete Screenshots der graphischen Oberfläche der verwendeten Simulations-Software LENS, reduziert auf die hier relevanten Input- und Output-Layer. Wie gehabt, zeigen wir in den einzelnen Teilabbildungen eine aufsteigende Aktivierungsrichtung: Die jeweils unterste Unit-Zeile stellt den sprachlichen Input dar, die folgende Unit-Zeile zeigt den sprachlichen Output (Wortvorhersage), und das Unit-Gitter zeigt den Output der visuellen Route. Die Abbildungen zeigen die Aktivierungen für die sukzessiven Positionen eines Beispielsatzes in Leserichtung von links nach rechts und von oben nach unten (entsprechend der Nummerierung von 1 bis 6). Ergänzend haben wir die korrespondierenden Wörter der Input-, Target- und tatsächlichen Output-Aktivierung vermerkt. Der im Beispiel dargebotene Satz lautet: *BOS links\_unter minus ist backslash EOS*. Verfolgen wir die hierfür entstehende Aktivungsverteilung:

1. Nach dem Input *BOS* bleiben der sprachliche und der visuelle Output weitgehend unspezifisch.
2. Dies ändert sich umgehend mit dem Input *links\_unter*: Im visuellen Output wird die passende Relation LINKSUNTER erfasst, wobei nur die für alle Objekte übereinstimmende mittlere Objekt-Unit stark aktiviert wird, während die umgebenden Units noch kein konkretes Objekt spezifizieren. Im sprachlichen Output werden korrekt Nomen für verschiedene Objekte vorhergesagt.
3. Der folgende Input *minus* führt im visuellen Output zu einer korrekten eindeutigen Spezifizierung des rechts oben gelegenen Objekts MINUS. Für das links unten gelegene Objekt etabliert sich der Slash. Im sprachlichen Output wird korrekt das Verb vorhergesagt.
4. Mit dem Verb *ist* ergeben sich im visuellen Output keine Veränderungen. Im sprachlichen Output werden der *slash* und der *backslash* vorhergesagt. Die Antizipation des Wortes *slash* als am stärksten aktiviertes Wort im sprachlichen Output korrespondiert zur Erwartung des Objekts SLASH im visuellen Output.
5. Erst mit dem erscheinen des letzten Inhalts-Wortes *backslash* wird im visuellen Output die korrekte Relation BACKSLASH LINKSUNTER MINUS aktiviert und die zuvor getroffene Annahme des Objekts SLASH revidiert. Im sprachlichen Output wird korrekt der Marker *EOS* vorhergesagt.
6. Mit dem abschließenden Input *EOS* wird das Startsymbol eines neuen Satzes *BOS* vorhergesagt, im visuellen Output wird die korrekte Relation noch etwas verstärkt.

Beispiel 10.3 verdeutlicht den inkrementellen Aufbau der zum Satz korrespondierenden visuellen Situation entsprechend der sukzessiven Erweiterung der gegebenen Informationen durch den eingehenden Satz. Unmittelbar mit der eingehenden Information wird die angenommene Situation angepasst. Nicht zur Verfügung stehende Informationen, etwa die Identität der Objekte vor ihrer expliziten Benennung, resultieren in unterspezifizierten Repräsentationen wie zum Beispiel auf Position (2) mit der Relation LINKSUNTER. Neben unterspezifizierten Repräsentationen werden auch willkürliche Annahmen getroffen, wie etwa auf Position (4) mit dem Objekt SLASH. Über den Grund für diese Auswahl kann nur spekuliert werden. Möglicherweise sind distributionale Eigenschaften dafür verant-

wortlich, etwa ein absolutes häufigeres Auftreten des Objekts SLASH im Trainings-Set oder ein relatives häufigeres Auftreten dieses Objekts in Kombination mit dem Objekt MINUS. Bei der sukzessiven Anpassung des visuellen Situationsmodells werden neue Informationen hinzugefügt, unterspezifizierte Repräsentationen konkretisiert und falsche Annahmen verworfen und korrigiert, so etwa bei der Änderung des Objekts SLASH zum Objekt BACKSLASH. Die starke Vorhersage von *slash* im sprachlichen Output auf Position (4) als zweites Nomen demonstriert die starke Verflochtenheit von sprachlicher und visueller Verarbeitung, durch die sich visuelle Erwartungen auch in der sprachlichen Antizipation widerspiegeln. Wortvorhersage wird nicht nur durch rein sprachinterne Wortfolge-Wahrscheinlichkeiten geleitet, sondern auch durch den visuellen (und in unserer Konzeption damit referenziell-semanticen) Kontext.<sup>15</sup>

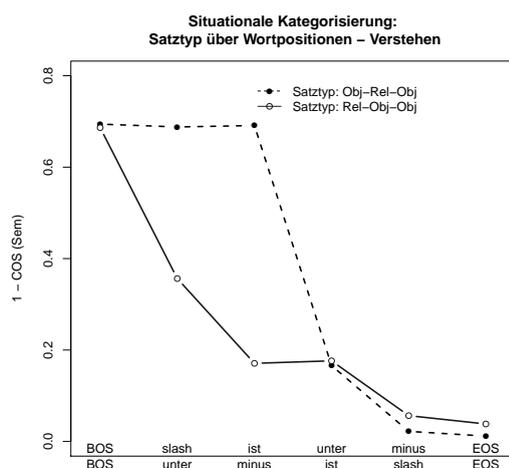


Abbildung 10.4.: SynSem I: Ergebnisse für die Performanz der visuellen Route bei der Muster-Assoziation (1-COS Sem) bei entkoppeltem Verstehen.

**Detailanalyse: Visuelle Route** Diagramm 10.4 stellt die Performanz bei der Konstruktion eines zum Satz korrespondierenden Situationsmodells dar. Wir vergleichen dafür den Fehlerwert 1-Cos Sem für den Gesamt-Vektor des visuellen Outputs für die beiden Satztypen ORO vs. ROO. Auf der x-Achse haben wir dafür die Wortfolgen der beiden Satztypen im Input der sprachlichen Route abgetragen. Im Diagramm wird ersichtlich, dass das Erscheinen des Adverbs den größten Einfluss auf die Performanz hat. Dieses erscheint im Satztyp ORO erst auf Position 4 und führt hier zu einer starken Abnahme des Fehlers. Während der Fehler zuvor konstant bleibt, nimmt er nach Erscheinen des Adverbs

<sup>15</sup>Beispiel 10.3 beschreibt die Aktivierungen bei Darbietung eines Satzes mit der Struktur ROO. Entsprechende Beispiele für die Satzstruktur ORO führen auf den frühen Positionen zu einer stark verteilten Aktivierung, die schwer zu interpretieren ist, aber darauf hindeutet, dass das erstgenannte Objekt auf verschiedenen Positionen des visuellen Outputs abgebildet wird. Mit Erscheinen des Adverbs erfolgt eine starke Anpassung der visuellen Output-Repräsentation, in der alle verfügbaren vorhergehenden Informationen integriert werden. Es erfolgen im Prinzip also die gleichen Update-Prozesse und die gleichen Verschränkungen visueller und sprachlicher Verarbeitung, wie sie für die Satzstruktur ORO beobachtet wurden.

kontinuierlich ab. Für beide Satztypen ist zu beobachten, dass der Fehler monoton fallend verläuft. Dabei verbleibt er jeweils auf der Position des Verbs konstant, da hier in unserer Mikrosprache und Mikrowelt keine inhaltlich-situationale Information übertragen wird. Die Diagramme in Abbildung 10.5 zergliedern die Bestandteile des Fehlers weiter, indem sie zwischen der Performanz bei der Objekt-Identifizierung und der Objekt-Lokalisierung unterscheiden. Diagramm 10.5(a) zeigt den Fehler bei der Objekt-Identifizierung (1-ObjCor): Für beide Satztypen wird nochmals ersichtlich, dass eine präzisere Abbildung der Objekte erst nach Kenntnis ihrer Relation möglich ist. Demzufolge fällt der Fehler im Satztyp ROO nach Nennung des ersten Objekts schneller, im Satztyp ORO schließt der Fehler erst nach Nennung des Adverbs auf das gleiche Niveau auf. Diagramm 10.5(b) zeigt den Fehler bei der Objekt-Lokalisierung (LocScore). Hier hängt der Fehler unmittelbar von der Nennung des Adverbs ab, nach der der Fehlerwert auf ein Minimum sinkt. In allen drei Diagrammen zeichnet sich ab, dass der Satztyp ORO auf den letzten Positionen die beste Performanz ermöglicht.

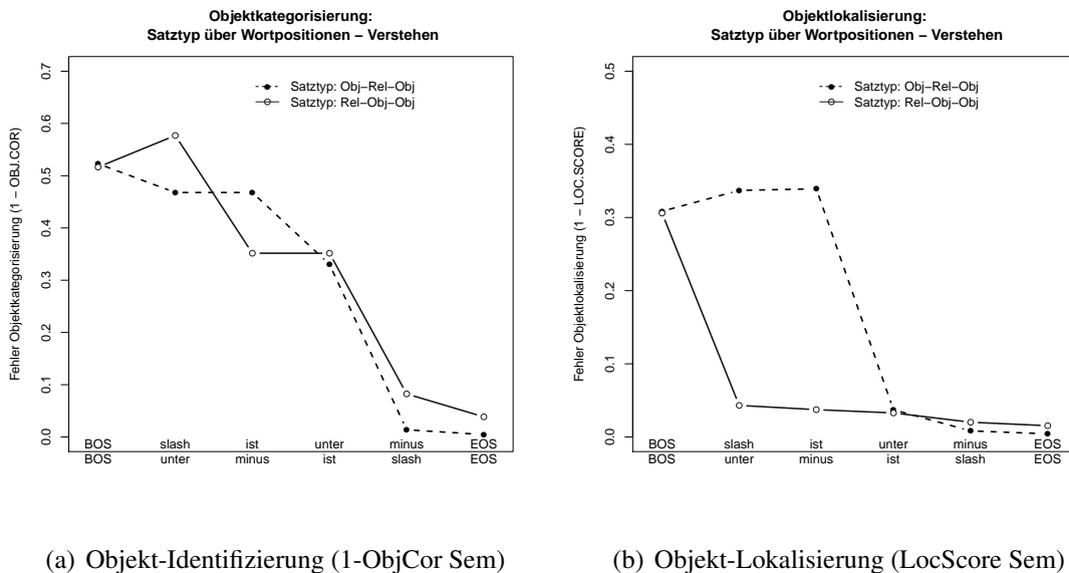
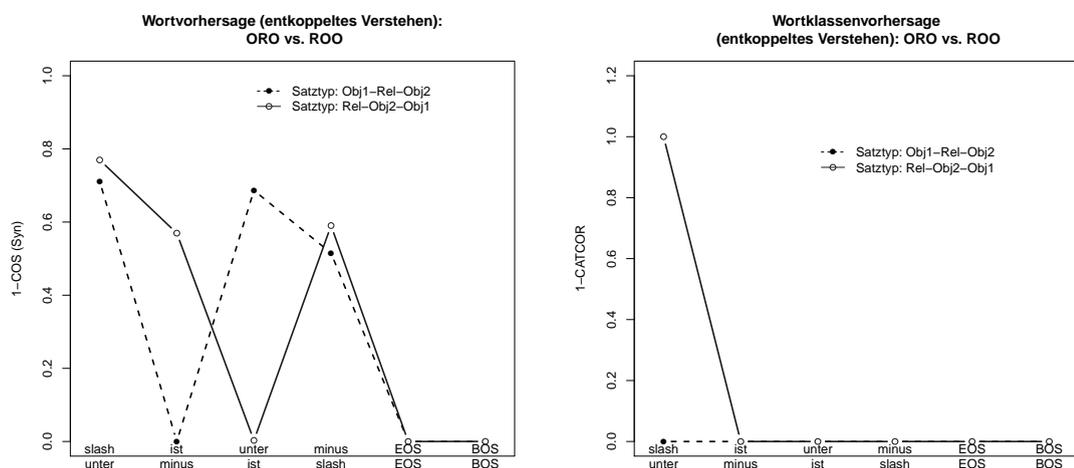


Abbildung 10.5.: SynSem I: Ergebnisse für Objekt-Identifizierung und -Lokalisierung im Test-Modus Verstehen. Grundlage ist der jeweils beste Vertreter beim Vergleich der Aktivierungsverteilungen mit möglichen Objekten und Lokalisierungen.

**Detailanalyse: Sprachliche Route** Die Diagramme in Abbildung 10.6 stellen die Performanz bei der Wortvorhersage dar. Hierbei tragen wir auf der x-Achse die Wortfolge der beiden Satztypen in Abhängigkeit des erwarteten Output-Worts ab.<sup>16</sup> Der in Diagramm 10.6(b) dargestellte Fehler der Vorhersage der Wortklasse ist und bleibt nach der ersten Position minimal. Dies ist nicht allzu verwunderlich, denn das Lexikon ist sehr beschränkt, es können nur wenige Wortklassen unterschieden werden, und vor allem ist unsere Mikrogrammatik sehr beschränkt. Der Fehler in Diagramm 10.6(a) basiert auf der Ähnlichkeit

<sup>16</sup>Anders als beim Fehler der visuellen Route, der vom aktuellen Input der sprachlichen Route bedingt wird, hängt der Fehler bei der sprachlichen Route maßgeblich davon ab, inwiefern das korrekte Wort oder die richtige Wortklasse vorhergesagt werden.

zwischen Output- und Target-Vektor (1-Cos Syn). Hier wird der oben erläuterte Sachverhalt ersichtlich, dass jeweils diejenigen Wortpositionen schlechter abschneiden, auf denen mehrere Wörter möglich sind – die Positionen mit Nomen und Adverbien. Da sich die Anzahl der Objekte und Relationen, auf die sprachlich verwiesen wird, im Test-Modus des entkoppelten Verstehens nicht anderweitig einschränken lässt, bleibt der Fehler auf den jeweiligen Positionen hoch. Auf den anderen Positionen (EOS, BOS, Verb) ist der Fehler hingegen minimal.



(a) Vorhersage Wörter (1-COS Syn)

(b) Vorhersage Wort-Kategorie (1-CatCor)

Abbildung 10.6.: SynSem I: Ergebnisse für Wortvorhersage mit Vergleich für die Ähnlichkeit von Aktivierungs- und Target-Vektor (1-COS Syn) gegenüber der Performanz bei der situationsunabhängigen Vorhersage der Wortklasse (1-CatCor).

**Zusammenfassung** Unseren Beispielen folgend wird (entkoppeltes) Sprachverstehen in unserem Modell als Konstruktion eines zur sprachlichen Beschreibung korrespondierenden Situationsmodells erfasst. Das Satzverstehen erfolgt inkrementell und in Übereinstimmung von visuellen und sprachlichen Informationen. Die ‚mentale‘ Simulation des Situationsmodells erfährt dabei ein sukzessives Updating, bei dem fehlende Information hinzugefügt, unterspezifizierte Information konkretisiert und falsche Annahmen korrigiert werden. Das im Aufbau befindliche Situationsmodell wird durch unterspezifizierte oder (stellvertretend) erwartete Repräsentations-Aspekte konsistent gehalten, um unvollständige oder unsichere Information zu überbrücken – im Sinne inkrementeller Verarbeitung leistet das Modell eine so vollständige Interpretation des Satzes wie möglich.

### 10.2.3.2. Ergebnisse 2: Situiertes Sprachverstehen

Wir analysieren für den Test-Modus des situierten Sprachverstehens, bei dem der Input beider Routen gegeben ist, zunächst wieder die globalen Kennwerte über alle Satzpositionen.

**Globale Kennwerte: Visuelle Route** Die getrennt ermittelten Werte für Relationen und Objekte betragen  $\bar{x}_{\text{LocScore Sem}} = 0.074$  ( $\sigma = 0.133$ ) und  $\bar{x}_{1\text{-ObjCor Sem}} = 0.16$  ( $\sigma = 0.249$ ). Sie sind damit wesentlich geringer als im Test-Modus mit entkoppeltem Verstehen. Dies zeigt sich wiederum in den Fehlerwerten für den gesamten visuellen Output:  $\bar{x}_{\text{EukDist Sem}} = 0.915$  ( $\sigma = 0.86$ ) und  $\bar{x}_{1\text{-Cos Sem}} = 0.161$  ( $\sigma = 0.231$ ). Vielleicht am deutlichsten zeigt sich der Unterschied für die Erfassung einer möglichen Relation, für die der Wert  $\bar{x}_{\text{PosRel Sem}} = 0.176$  ( $\sigma = 0.381$ ) ermittelt wurde. Demzufolge werden bereits auf frühen Positionen nur mögliche Objekt-Relationen konstruiert. Auch die gemittelten Fehlerwerte für die abschließende Position des EOS-Markers zeigen eine Verbesserung im Vergleich zum entkoppelten Verstehen:  $\bar{x}_{\text{LocScore Sem}} = 0.001$  ( $\sigma = 0.013$ ),  $\bar{x}_{1\text{-ObjCor Sem}} = 0.002$  ( $\sigma = 0.027$ ),  $\bar{x}_{\text{EukDist Sem}} = 0.076$  ( $\sigma = 0.172$ ),  $\bar{x}_{1\text{-Cos Sem}} = 0.003$  ( $\sigma = 0.025$ ),  $\bar{x}_{\text{PosRel Sem}} = 0.003$  ( $\sigma = 0.055$ ). Der positive Effekt scheint sich also bis zum jeweiligen Satzende auf die Performanz auszuwirken. Vor allem auf den frühen Positionen ist dies natürlich nicht unerwartet, denn schließlich verfügt das Netzwerk bereits von Beginn an über die meisten relevanten Informationen. Einzig die Information über die sprachlich ausgewählte Objekt-Relation (im Falle von drei Input-Objekten) kann nur über den sprachlichen Input erschlossen werden.

**Globale Kennwerte: Sprachliche Route** Beim Vergleich der Wortvorhersage zeigt sich für den Fehler der kategorialen Vorhersage ein identischer Wert wie beim entkoppelten Verstehen:  $\bar{x}_{1\text{-CatCor Syn}} = 0.081$  ( $\sigma = 0.272$ ). Auf die rein morphosyntaktische Verarbeitung der Wortfolge-Wahrscheinlichkeit scheint sich die Situiertheit also nicht auszuwirken; demnach erfolgt diese unabhängig vom visuellen Input bzw. der semantischen Verarbeitung. Diese Aspekte wirken sich aber aus, wenn es darum geht, die Vorhersage semantisch zu präzisieren: Die positionsübergreifenden auf Distanz und Ähnlichkeit basierenden Fehlerwerte sind mit  $\bar{x}_{\text{EukDist Syn}} = 0.385$  ( $\sigma = 0.424$ ) und  $\bar{x}_{1\text{-Cos Syn}} = 0.222$  ( $\sigma = 0.3$ ) deutlich geringer als beim entkoppelten Verstehen. Auf der abschließenden Position hingegen fallen die Fehlerwerte wieder weitgehend gleich aus mit  $\bar{x}_{1\text{-CatCor Syn}} = 0.001$  ( $\sigma = 0.032$ ),  $\bar{x}_{\text{EukDist Syn}} = 0.004$  ( $\sigma = 0.03$ ) und  $\bar{x}_{1\text{-Cos Syn}} < 0.001$  ( $\sigma = 0.015$ ). Die reduzierten positionsübergreifenden Fehlerwerte EukDist und 1-Cos deuten auf eine Reduzierung der in Betracht gezogenen Wörter auf den Objekt- und Adverb-Positionen hin. Der visuelle Input scheint dafür genutzt zu werden, eine genauere Wortvorhersage zu treffen.

**Beispiel: Inferenz bei situiertem Verstehen** Wie wirkt sich die Verschränkung der visuellen und sprachlichen Verarbeitung mit gegenseitiger Information und Beschränkung konkret aus? Wir beginnen mit einem Beispiel, das den starken gegenseitigen Einfluss besonders eindrücklich demonstriert. Abbildung 10.7 zeigt die Aktivierungs-Details bei situiertem Verstehen für den Satz *BOS über pipe ist minus EOS*. Im visuellen Input sind drei mögliche Relationen gegeben, die beschrieben werden könnten (vgl. untere Unit-Gitter der Teilabbildungen): MINUS ÜBER PIPE, MINUS LINKSÜBER SLASH und PIPE LINKSVON SLASH. Die sukzessive Darbietung des Satzes führt zu folgenden Output-Aktivierungen:

1. Mit dem Input *BOS* finden wir wieder die unspezifische Aktivierungsverteilung im sprachlichen Output. Im Unterschied zum entkoppelten Verstehen finden wir aber beim situierten Verstehen für den visuellen Output ein interessantes Aktivierungs-

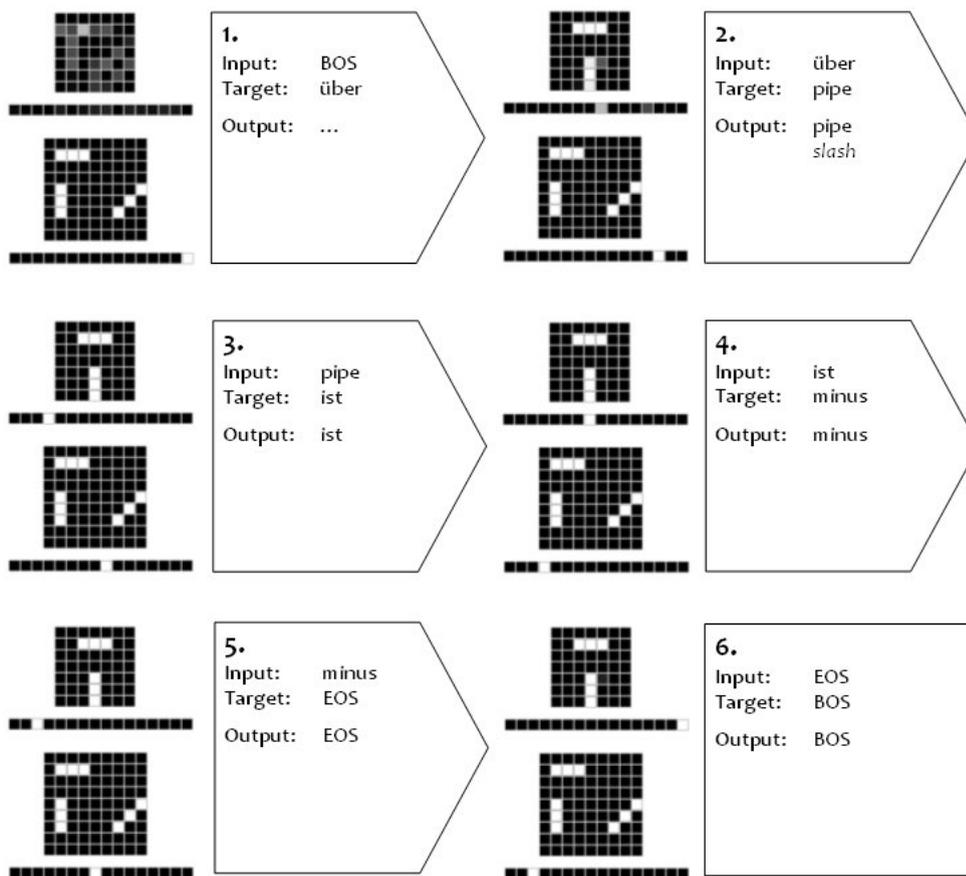


Abbildung 10.7.: SynSem I: Beispiel zum situierten Verstehen eines Satzes und zur gegenseitigen Nutzung der Informationen des anderen Input-Systems für Inferenzen.

Muster: Auf dem Unit-Gitter findet sich eine vage Abbildung aller drei möglichen Relationen.<sup>17</sup>

2. Der folgende Input *über* löst jeglichen weiteren Disambiguierungsbedarf: Mit der sprachlichen Benennung der relevanten Relation durch das Adverb können die Relationen MINUS LINKSÜBER SLASH und PIPE LINKSVON SLASH ausgeschlossen werden – die einzige Relation des visuellen Inputs, die über das Merkmal ÜBER verfügt, ist MINUS ÜBER PIPE. Der visuelle Output bildet fast ausschließlich diese Relation ab. Ganz leicht deutet sich für das untere Objekt der in der Situation ebenfalls gegebene SLASH an. Dies deckt sich mit der sprachlichen Vorhersage auf dieser Position: Neben der dominanten Aktivierung von *pipe* wird marginal auch *slash* aktiviert.
3. Mit dem folgenden Input *pipe* klärt sich die Abbildung im visuellen Output endgültig, auf den Folgepositionen ergibt sich hier keine Veränderung mehr.
4. Die Vorhersagen im sprachliche Output sind ab einschließlich Position 3 auf das jeweilige einzelne Element beschränkt, das im gegebenen visuellen und sprachlichen Kontext möglich ist. Die Disambiguierung der Situation mit der Identifizierung der relevanten Relation wirkt sich also gleichermaßen auf die Abbildung des visuellen Outputs und auf die Vorhersage des sprachlichen Outputs aus.
5. Ereignislose korrekte Weiterführung des Satzes.
6. Ereignisloser korrekter Abschluss des Satzes.

Die Performanz auf der ersten Position ist vergleichbar mit der Konjunktion und Disjunktion der Belegung des Repräsentationsraums im DSS-Modell (S. L. Frank et al., 2003b; S. L. Frank & Haselager, 2006): Je weniger Informationen gegeben sind oder je allgemeiner diese sind, desto größer ist der Anteil des Repräsentationsraums, der bei der Aktivierung ausgeschöpft wird. Je mehr Information durch Konjunktion und gegenseitige Beschränkung gegeben ist, desto geringer fällt der Umfang des beanspruchten Repräsentationsraums aus, die Aktivierung verteilt sich auf eine geringere Anzahl relevanter Units. Dies führt durch die gegenseitigen Beschränkungen der sprachlichen und visuellen Informationen (gegebenes Adverb und mögliche Relationen) ab der zweiten Position zur demonstrierten Reduzierung der Aktivierung. Diese Reduzierung geht weit über das beim entkoppelten Verstehen gezeigte Verhalten hinaus: Es wird nicht nur eine ÜBER-Relation aktiviert, vielmehr werden bereits die relevanten Objekte differenziert. Dies entspricht einer situationalen Inferenz: Die weiteren Verhältnisse der gesamten Situation werden aus einem Teilaspekt geschlossen. Das gezeigte Verhalten kann des Weiteren als Lenkung der visuellen Aufmerksamkeit (covert attention) interpretiert werden, wie es im Modell CIA-Net (Mayberry et al., 2005, 2009) simuliert wird. Durch den sprachlichen Input wird die visuelle Aufmerksamkeit unmittelbar auf den relevanten Aspekt der Situation gelenkt. Im Unterschied zu CIA-Net erfolgt die Lenkung der Aufmerksamkeit aber intrinsisch durch die Reduzierung bzw. Prototypisierung der räumlichen Objekt-Konstellation und wird nicht durch einen explizit implementierten Mechanismus herbeigeführt. Diese Simulation visueller Aufmerksamkeit entspricht damit mehr den Annahmen von Altmann und Mirkovic (2009). Mit dem gegebenen Beispiel wird auch demonstriert, dass die beiden

<sup>17</sup>Mit abnehmender Deutlichkeit bzw. Vollständigkeit der Objekte: MINUS LINKSÜBER SLASH, PIPE LINKSVON SLASH, MINUS ÜBER PIPE).

Routen auf eine Weise integriert sind, die mit den gegenseitigen Beschränkungen eine perfekte Antizipation für beide Aufgaben ermöglicht. Im Unterschied zum entkoppelten Verstehen werden deshalb bei der Wortvorhersage nur die Wörter aktiviert, die neben ihrer Wortfolge-Wahrscheinlichkeit auch die semantischen Beschränkungen durch die gegebene Situation erfüllen.

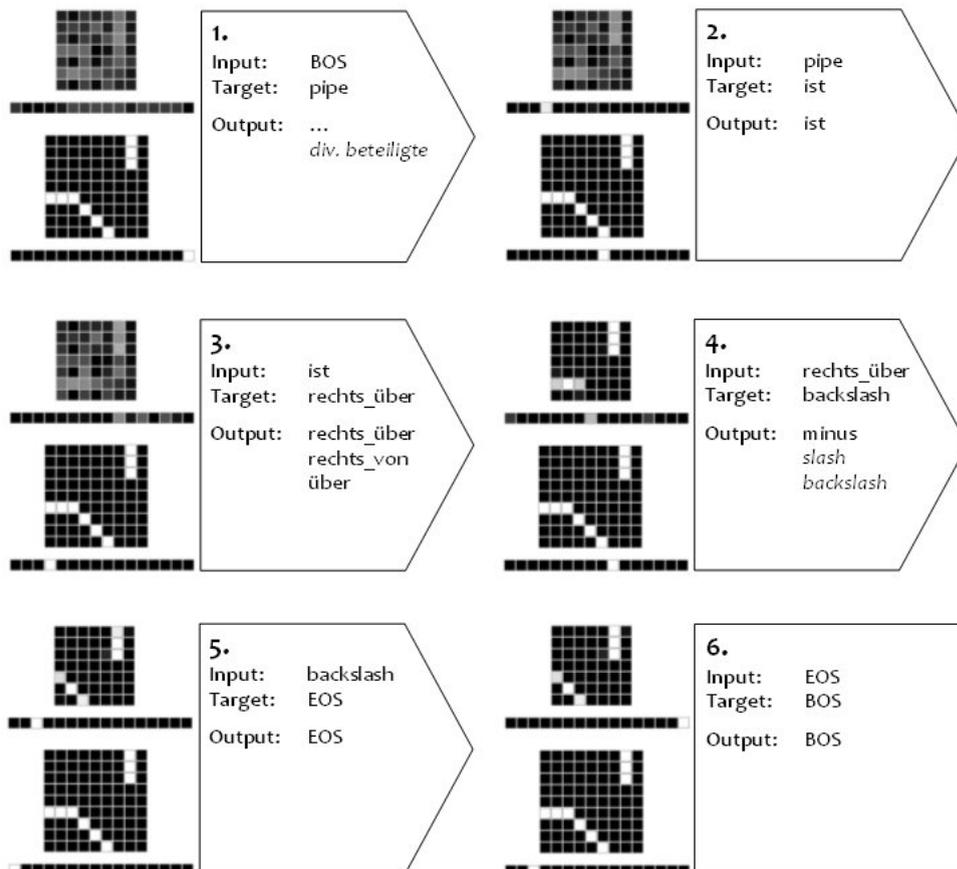


Abbildung 10.8.: SynSem I: Beispiel zum situierten Verstehen eines Satzes und zum inkrementellen Aufbau des Objekt-Relation-Modells bei ambigen visuellen Informationen.

**Beispiel: Situiertes Verstehen und Ambiguität** Bevor wir zu den Diagrammen übergehen und weitere Performanz-Aspekte detaillierter besprechen, demonstrieren wir an einem zweiten Beispiel die Performanz für einen Satz und einen visuellen Kontext, bei dem eine frühzeitige Disambiguierung nicht möglich ist. Abbildung 10.8 zeigt die Aktivierungen für den Satz *BOS pipe ist rechts\_über backslash EOS*. Der visuelle Kontext umfasst mit drei Objekten wiederum drei mögliche Relationen: PIPE RECHTSÜBER BACKSLASH, PIPE RECHTSÜBER MINUS und MINUS LINKSÜBER BACKSLASH. Die Ziel-Relation ist PIPE RECHTSÜBER BACKSLASH. Dadurch, dass zwei RECHTSÜBER-Relationen gegeben sind, lässt sich die Ziel-Relation schwerer identifizieren. Die sukzessive Darbietung des Satzes führt unter diesen Voraussetzungen zu folgenden Output-Aktivierungen:

1. Nach dem Input *BOS* bleibt der Output beider Routen weitgehend unspezifisch, allerdings deutet sich im visuellen Output eine Aktivierung der Teil-Relation PIPE RECHTSÜBER an.
2. Mit dem Input *pipe* verdeutlicht sich im visuellen Output die Aktivierung des Objekts PIPE. Zudem deutet sich die Aktivierung des Objekts MINUS an, wobei dieses eine Position für die Relation LINKSUNTER und etwas geringfügiger für die Relation LINKSVON einnimmt. Insgesamt bleibt die Aktivierung aber noch stark verteilt und vage. Im sprachlichen Output wird korrekt *ist* vorhergesagt.
3. Nach dem Input *ist*, mit dem keine neue inhaltlich-situationale Information hinzugefügt wird, bleibt die Aktivierung im visuellen Output stark verteilt und vage. Im sprachlichen Output wird am stärksten die für das bereits benannte Objekt PIPE das Wort für die Relation RECHTSÜBER vorhergesagt. Etwas geringfügiger werden *rechts\_von* und *über* vorhergesagt.
4. Mit dem Input *rechts\_über* wird die Relation im visuellen Output schließlich konkretisiert als PIPE RECHTSÜBER MINUS. Dementsprechend wird im sprachlichen Output das passende Nomen zur antizipierten Relation *minus* vorhergesagt. Deutlich schwächer werden auch die Wörter *slash* und *backslash* aktiviert.
5. Erst auf der fünften Position mit der Benennung des zweiten Objekts mit dem Nomen *backslash* wird die eigentlich intendierte Relation abgebildet und die Abbildung im visuellen Output entsprechend korrigiert. Korrekt wird der *EOS* vorhergesagt.
6. Abschließend wird korrekt der neue Satzanfang mit *BOS* vorhergesagt.

Zunächst einmal zeigt dieses Beispiel den Umgang des Modells mit Ambiguität und die inkrementelle Verarbeitung beim situierten Sprachverstehen. Im visuellen Output äußert sich dies in der Konstruktion unterspezifizierter Repräsentationen, bei der Wortvorhersage in der Aktivierung verschiedener semantisch möglicher Alternativen. Gerade dies zeigt sich im Beispiel: Das Modell berücksichtigt bei den Wortvorhersagen mehrere (weitgehend) zum visuellen Input korrespondierende semantische Constraints – hier ersichtlich auf den Positionen 3 und 4.<sup>18</sup> Hierbei ist zu erläutern, warum die Vorhersage von *rechts\_von* und *über* auf Position 3 nicht ganz unbegründet ist: Der Algorithmus, der bei der Erzeugung der Stimuli über eine Winkel-Funktion die prototypischen Relationen bestimmt, verschiebt mit zunehmender Distanz zweier Objekte die Wahrscheinlichkeit dafür, dass die Objekte als diagonal oder aber als horizontal/vertikal zueinander klassifiziert werden. Durch den vergleichsweise großen Abstand im Beispiel könnte das Objekt PIPE grenzwertig in den Relationen RECHTSVON MINUS und ÜBER BACKSLASH erkannt werden. Hierfür wären die Vorhersagen von *rechts\_von* sowie *über* korrekt. Da die damit beschriebenen Relationen aber eher Grenzfälle wären, werden sie dementsprechend geringfügig in der Wortvorhersage aktiviert. Die Aktivierungsverteilung ist sogar vielmehr ein weiterer Beleg dafür, dass das Modell Restriktionen mehrerer möglicher Interpretationen parallel berücksichtigen und zudem nach ihrer Wahrscheinlichkeit gewichten kann.

<sup>18</sup>Einzige Ausnahme ist die minimale Vorhersage des Nomens *slash*, dessen Referenz-Objekt in der gegebenen Situation nicht vorkommt.

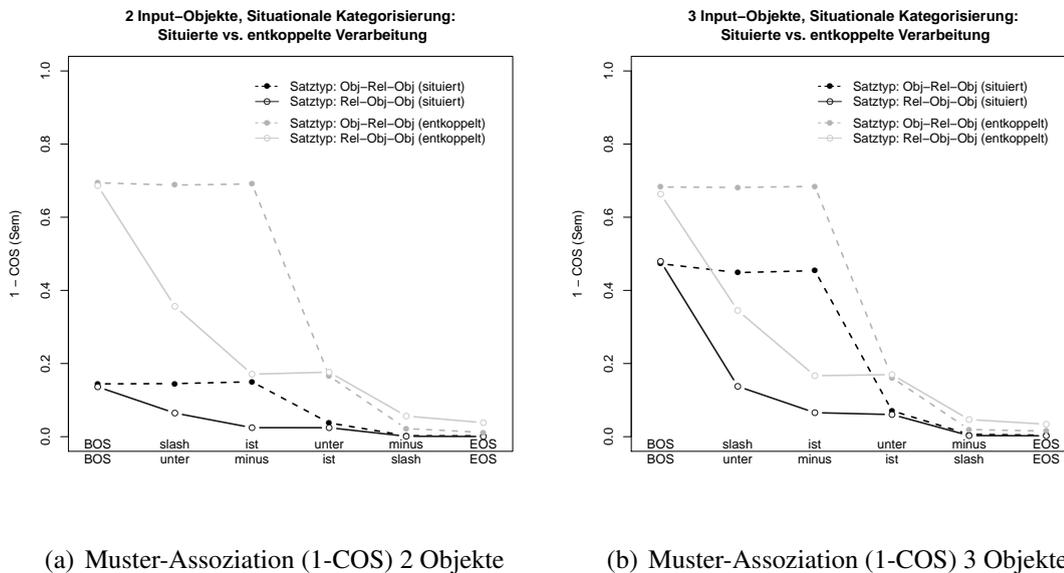
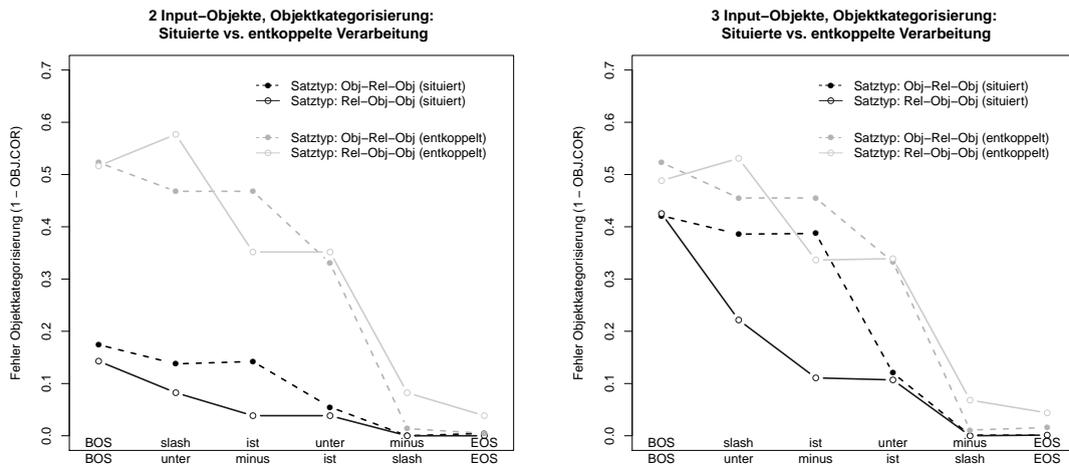


Abbildung 10.9.: SynSem I: Ergebnisse für die Muster-Assoziation im Test-Modus situierte Sprachverarbeitung (1-COS Sem). Innerhalb der Diagramme erfolgt der Vergleich des situierten Modus zum entkoppelten Verstehen. Über die Diagramme hinweg wird zwischen den Input-Situationen mit zwei und drei Objekten unterschieden.

**Detailanalyse: Visuelle Route** Abbildung 10.9 enthält Diagramme zur Performanz der visuellen Route. Auf der x-Achse werden wieder die Input-Wörter der beiden Satztypen abgetragen, das dargestellte Fehlermaß ist 1-Cos Sem. Wir vergleichen damit neben den beiden Satztypen ORO und ROO zum einen den Test-Modus der situierten Verarbeitung mit dem entkoppelten Satzverstehen, um die Vorteile der Nutzung der Informationen des situationalen Kontexts darzustellen. Zum anderen vergleichen wir über die beiden Diagramme hinweg den Einfluss der Anzahl der gegebenen Input-Objekte. Zunächst ist zu beobachten, dass der Fehler für die Bedingung mit situiertem Verstehen in beiden Objekt-Bedingungen von Beginn an geringer ausfällt als in der Bedingung mit entkoppeltem Verstehen. Folglich können die Informationen sehr gut für die Konkretisierung des Situationsmodells genutzt werden. Dabei zeigt sich auch wieder die monoton fallende Verbesserung des Fehlers, bei der für die situierte Verarbeitung früher eine vollständige Fehlerminimierung erreicht wird. Die Bedingungen mit zwei gegenüber drei Input-Objekten unterscheiden sich vor allem durch einen wesentlich niedrigeren Ausgangsfehler auf Satzposition 1. In der Bedingung mit drei Objekten bleibt der erhöhte Fehler im Satztyp ORO bis zum Erscheinen des Adverbs erhalten. Der Fehler gleicht sich erst auf Position 5 für beide Objekt-Bedingungen aneinander an. Der höhere Fehler in der Bedingung mit drei Objekten ist dadurch bedingt, dass bis zum Erscheinen des Adverbs nicht entschieden werden kann, welche der drei möglichen Relationen als Gegenstand der Beschreibung intendiert ist. Auch bei nur zwei Objekten ist der Fehler vor Erscheinen des Adverbs nicht minimal, auch wenn er in dieser Bedingung sehr viel geringer ausfällt. Allerdings wäre zu erwarten gewesen, dass bei nur zwei Input-Objekten von Beginn an eine korrekte Output-Aktivierung möglich ist. Der sprachliche Input scheint hier eine stützende bzw. konkretisierende Funktion bei der

Kategorisierung der Situation zu erfüllen, die über die bloße Wahrnehmung der Situation hinausgeht.

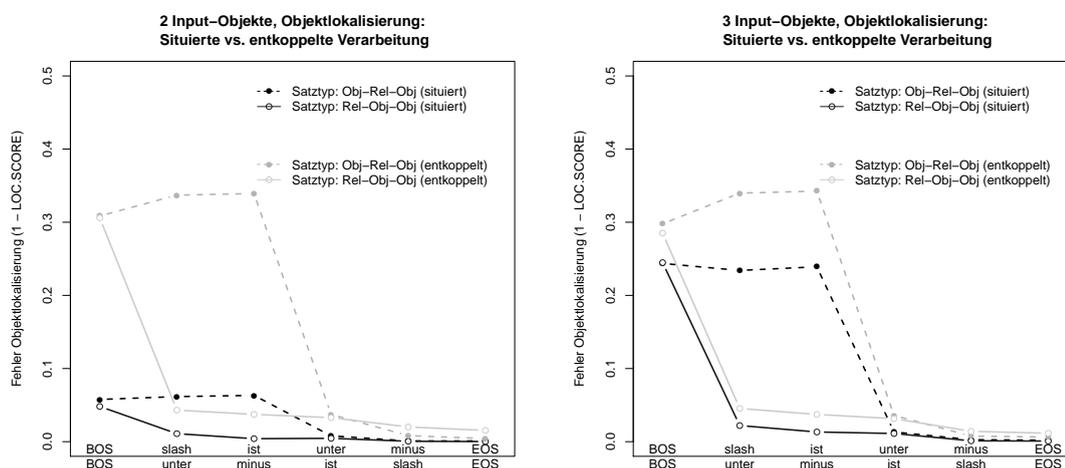


(a) Objekt-Identifizierung (1-ObjCor) 2 Objekte (b) Objekt-Identifizierung (1-ObjCor) 3 Objekte

Abbildung 10.10.: SynSem I: Ergebnisse für die Objekt-Identifizierung (Fehlermaß 1-ObjCor) im Test-Modus situierte Sprachverarbeitung für zwei (linke Abb.) oder drei Objekte (rechte Abb.) im Vergleich zur Performanz bei entkoppeltem Verstehen.

Die beiden Abbildungen 10.10 und 10.11 zeigen wiederum ergänzend die Performanz bei der Objekt-Identifizierung sowie bei der Objekt-Lokalisierung. Wir stellen dabei ebenfalls die Bedingungen mit zwei und drei Objekten sowie die Test-Modi mit entkoppeltem und situiertem Sprachverstehen gegenüber. Sowohl im Fehler der Identifizierung als auch der Lokalisierung spiegeln sich die Eigenschaften wider, die wir bereits für den ähnlichkeitsbasierten Vergleich des generierten Situationsmodells mit 1-Cos Sem beschrieben haben. Dabei zeigt sich auch für beide Aufgaben die Bedeutung der Nennung des Adverbs, durch dessen Auswahlfunktion der intendierten Relation beide Aspekte deutlich besser konkretisiert werden können. Besonders im Vergleich der Objekt-Identifizierung bei situiertem gegenüber entkoppeltem Verstehen bei drei Objekten 10.10(b) zeigt sich das relativ hohe Gewicht der Information des Adverbs, während dieses beim entkoppelten Verstehen keinen (bei ROO) oder nur geringen (bei ORO) Einfluss nimmt. Der Einfluss der Nennung des Adverbs wirkt sich auch in der Bedingung mit nur zwei Objekten aus, hier allerdings deutlich geringfügiger. Anhand des Lokalisations-Fehlers wird ersichtlich, dass die Verortung der Objekte hier in beiden Satzbedingungen bereits weitgehend abgeschlossen ist und dementsprechend eine Verbesserung ihrer Identifizierung fast nur noch durch die Benennung der Objekte erreicht wird.

**Detailanalyse: Sprachliche Route** Für die Performanz bei der Wortvorhersage verzichten wir auf eine Abbildung des Fehlerverlaufs der Wortklassen-Vorhersage; wie bereits beim entkoppelten Verstehen bleibt der Fehler hier nach der ersten Position minimal. Abbildung 10.12 zeigt den ähnlichkeitsbasierten Vergleich von Output und Target für die

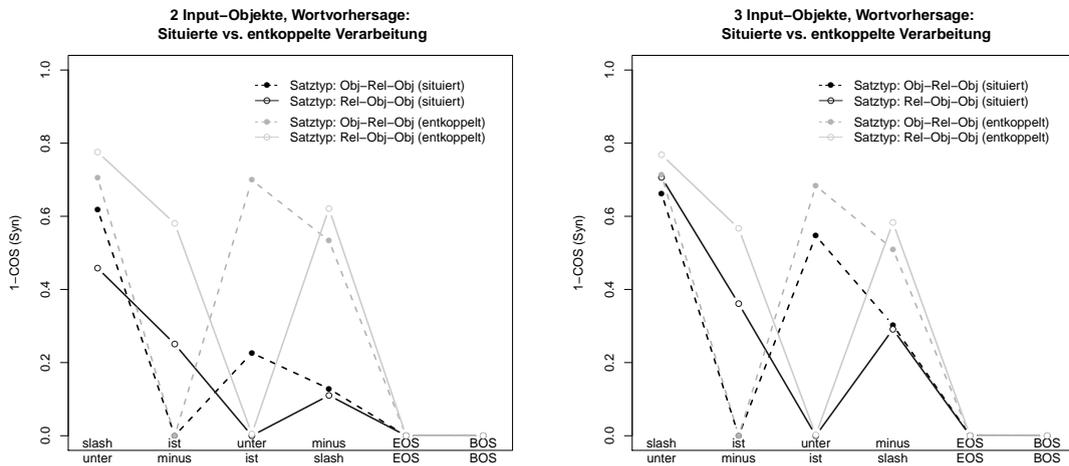


(a) Objekt-Lokalisierung (LocScore) 2 Objekte      (b) Objekt-Lokalisierung (LocScore) 3 Objekte

Abbildung 10.11.: SynSem I: Ergebnisse für die Objekt-Lokalisierung (Fehlermaß LocScore) im Test-Modus situierte Sprachverarbeitung für zwei (linke Abb.) oder drei Objekte (rechte Abb.) im Vergleich zur Performanz bei entkoppeltem Verstehen.

Wortvorhersage. Wir erfassen dabei wieder die Satztypen ORO und ROO, den Test-Modus (situert vs. entkoppelt) sowie über die Diagramme hinweg die Anzahl der im Input erscheinenden Objekte (2 vs. 3). Während das Maß 1-Cos Syn für das entkoppelte Satzverstehen nur bedingt sinnvoll war, da der Fehler in erster Linie durch die unterschiedliche Anzahl (korrekt) aktivierter Elemente bedingt war, zeigt sich nun bei der situierten Verarbeitung der eigentliche Nutzen des Maßes. Zum einen fallen innerhalb der jeweiligen Satzbedingungen fast alle Fehler geringer (und kein Fehler größer) aus als beim entkoppelten Verstehen. Dies zeigt, dass die Situation im Sinne eines Priming-Effekts die Aktivierung derjenigen Wörter verstärkt, deren Referenz-Objekte und -Relationen im situationalen Kontext gegeben sind. Da die möglichen Referenzen in der Bedingung mit nur zwei Objekten noch stärker beschränkt sind, wirkt sich der Effekt hier stärker aus. Zudem zeigt sich eine Interaktion zwischen sprachlichem und visuellem Input: Die bereits benannten Objekte der Situation können nicht nochmals genannt werden, ebenfalls können durch die Benennung der Relation Objekt-Benennungen ausgeschlossen werden. Dadurch ergibt sich eine stärkere Verringerung des Fehlers auf Folgepositionen, der zum Beispiel in allen Bedingungen (Objekt-Anzahl und Satztyp) auf Satzposition 4 bei der Benennung des zweiten Objekts ersichtlich wird.

**Zusammenfassung** Die Ergebnisse für diesen Test-Modus zeigen, dass SRNs wesentlich mehr als nur Wortklassen-Vorhersagen im Sinne des Part-Of-Speech-Tagging (vgl. Steedman, 1999) leisten können. Durch die integrativen Leistungsmerkmale können die Einflüsse verschiedener sprachlicher und nicht-sprachlicher Kontexte im Sinne referenzieller Verarbeitung auf die sprachliche Performanz einwirken. Dabei werden nicht alle Wörter aktiviert, die sich mit der entsprechend zur Position passenden Kollokation auszeichnen, vielmehr ergibt sich eine referenziell-semantische Spezifizierung der lexikalischen Ak-



(a) Vorhersage Wörter (1-COS Syn) 2 Objekte

(b) Vorhersage Wörter (1-COS Syn) 3 Objekte

Abbildung 10.12.: SynSem I: Ergebnisse für die Wortvorhersage bei situierter Sprachverarbeitung anhand des Vergleichs der Ähnlichkeit des Aktivierungs- und Target-Vektors (1-COS Syn). Innerhalb der Diagramme erfolgt wiederum der Vergleich des situierten Modus zum entkoppelten Verstehen, über die beiden Diagramme wird die Performanz für zwei gegenüber drei Input-Objekten verglichen.

tivierung – es werden nicht mehr nur Types, sondern auch reduzierend einzelne Tokens vorhergesagt. Wir haben für das Modell AnaSRN festgestellt, dass sich SRNs hier am ehesten als Modell früher syntaktischer Verarbeitung bzw. als initialer Filter der Interpretation beschreiben lassen (vgl. Kap. 6.6). Die Verarbeitungseigenschaften der sprachlichen Route des Modells SynSem rücken SRNs hingegen deutlich näher zur Ebene der interpretativen Verarbeitung, denn im Verlauf der Verarbeitung werden sprachliche Sequenzen nachvollzogen, die lexikalisch-syntaktisch sowie referenziell-semantisch spezifiziert sind. Die Verarbeitung der visuellen Route zeigt wieder die monoton-inkrementelle Verarbeitung. Dabei zeigt sich, dass nicht nur sprachliche, sondern auch nicht-sprachliche Informationen in diese inkrementelle Verarbeitung eingreifen. Das Situationsmodell wird unmittelbar mit der hinzukommenden Information so umfassend wie möglich angepasst. Dabei werden alle Informationen berücksichtigt, auch Informationen bzw. Beschränkungen des visuellen Kontexts werden unmittelbar integriert. Die enge modalitätsübergreifende Korrespondenz wirkt sich auch auf die sprachliche und visuelle Antizipation aus. Wird im visuellen Output ein Objekt verstärkt angenommen, so wird verstärkt das korrespondierende Nomen vorhergesagt – selbst wenn es sich im weiteren Verlauf des Satzes als falsch herausstellt und das abgebildete Objekt revidiert werden muss. Wir finden also für die visuelle und gleichermaßen die sprachliche Verarbeitung Leistungsmerkmale, die im Sinne von kontextabhängigen Inferenzen bzw. als Lenkung visueller Aufmerksamkeit sowie als modalitätsübergreifende Priming-Effekte beschrieben werden können. Diese Effekte wurden nicht explizit implementiert wie etwa im Modell CIA-Net (Mayberry et al., 2009), sondern entstehen als Nebenprodukt der integrativen Verarbeitung, der selbstorganisierten Aufgabenkoordination sowie der prototypischen Reduzierung der Target-Repräsentation.

### 10.2.3.3. Ergebnisse 3: Produktion

Wir analysieren die globalen Fehlerwerte in der 15. Epoche über alle Satzpositionen hinweg im Test-Modus Sprachproduktion, in dem ausschließlich visueller Input gegeben ist und das Netzwerk um eine Feedback-Schleife an der sprachlichen Route und eine Winner-takes-all-Funktion erweitert wird.

**Globale Kennwerte: Visuelle Route** Die Fehlerwerte des Outputs der visuellen Route sind im Vergleich zu den anderen Test-Modi sehr hoch. Darüber hinaus bleiben sie zudem bis zur jeweils letzten Satzposition hoch.<sup>19</sup> Die Werte sind allerdings direkt davon abhängig, ob im Input zwei oder drei Objekte dargeboten werden. So sind die Kennwerte bei nur zwei Input-Objekten wesentlich besser als bei drei Input-Objekten:  $\bar{x}_{\text{LocScore Sem}} = 0.077$  ( $\sigma = 0.127$ ),  $\bar{x}_{1\text{-ObjCor Sem}} = 0.218$  ( $\sigma = 0.255$ ),  $\bar{x}_{\text{EukDist Sem}} = 1.364$  ( $\sigma = 0.786$ ),  $\bar{x}_{1\text{-Cos Sem}} = 0.228$  ( $\sigma = 0.232$ ) und  $\bar{x}_{\text{PosRel Sem}} = 0.461$  ( $\sigma = 0.499$ ). Im Vergleich hierzu die Werte bei drei Input-Objekten:  $\bar{x}_{\text{LocScore Sem}} = 0.327$  ( $\sigma = 0.23$ ),  $\bar{x}_{1\text{-ObjCor Sem}} = 0.466$  ( $\sigma = 0.285$ ),  $\bar{x}_{\text{EukDist Sem}} = 2.385$  ( $\sigma = 0.695$ ),  $\bar{x}_{1\text{-Cos Sem}} = 0.628$  ( $\sigma = 0.306$ ) und  $\bar{x}_{\text{PosRel Sem}} = 0.612$  ( $\sigma = 0.487$ ). Die schlechteren Werte bei drei Objekten sind direkt darauf zurückzuführen, dass für die Identifizierung der ausgewählten Objekt-Relation die sprachliche Beschreibung benötigt wird, während bei nur zwei Objekten bereits alle Informationen für die Identifikation gegeben sind.

**Globale Kennwerte: Sprachliche Route** Für die Performanz bei der Wortvorhersage spielt die Unterscheidung zwischen zwei und drei Objekten hingegen keine Rolle, die Fehlerwerte sind in diesen beiden Bedingungen nahezu identisch. Im Vergleich zu den anderen Test-Modi sind die Fehlerwerte deutlich erhöht:  $\bar{x}_{1\text{-CatCor Syn}} = 0.275$  ( $\sigma = 0.447$ ) und  $\bar{x}_{\text{EukDist Syn}} = 0.735$  ( $\sigma = 0.275$ ).<sup>20</sup> Für einzelne Satzpositionen sind die Fehlerwerte der Wortklassen-Vorhersage besser, speziell für den EOS-Marker sowie für die ersten beiden Satzpositionen.<sup>21</sup> Insgesamt betrachtet sind die Fehlerwerte aber schlechter als erwartet und deuten darauf hin, dass unsere Erwartungen besonders im Falle der Abfolge bestimmter Wortklassen im Sinne einer syntaktisch wohlgeformten Sequenz nur bedingt erfüllt wurden.

**Beispiel: Beschreibende Sprachproduktion** Wir ziehen auch für den Test-Modus der Satzproduktion unter situational-semantic Beschränkungen ein Beispiel heran. Abbildung 10.13 zeigt ein Beispiel mit erfolgreicher Performanz. Dies soll allerdings nicht

<sup>19</sup>Die positionsübergreifenden Werte für den visuellen Output unabhängig von der Anzahl der Input-Objekte betragen  $\bar{x}_{\text{LocScore Sem}} = 0.277$  ( $\sigma = 0.236$ ),  $\bar{x}_{1\text{-ObjCor Sem}} = 0.416$  ( $\sigma = 0.297$ ),  $\bar{x}_{\text{EukDist Sem}} = 2.181$  ( $\sigma = 0.823$ ),  $\bar{x}_{1\text{-Cos Sem}} = 0.548$  ( $\sigma = 0.334$ ) und  $\bar{x}_{\text{PosRel Sem}} = 0.582$  ( $\sigma = 0.493$ ). Auf der satzfinalen Position beträgt der Fehler immer noch  $\bar{x}_{\text{LocScore Sem}} = 0.301$  ( $\sigma = 0.251$ ),  $\bar{x}_{1\text{-ObjCor Sem}} = 0.433$  ( $\sigma = 0.303$ ),  $\bar{x}_{\text{EukDist Sem}} = 2.328$  ( $\sigma = 0.889$ ),  $\bar{x}_{1\text{-Cos Sem}} = 0.602$  ( $\sigma = 0.34$ ) und  $\bar{x}_{\text{PosRel Sem}} = 0.685$  ( $\sigma = 0.465$ ).

<sup>20</sup>Die Berechnung des Fehlerwerts 1-Cos Syn war aufgrund der Datenlage nicht möglich.

<sup>21</sup>Folgende Fehlerwerte bestehen abhängig von der Satzposition: Satzanfangs-Marker  $\bar{x}_{1\text{-CatCor Syn}} = 0.047$  ( $\sigma = 0.212$ ), Pos. 1  $\bar{x}_{1\text{-CatCor Syn}} = 0.017$  ( $\sigma = 0.129$ ), Pos. 2  $\bar{x}_{1\text{-CatCor Syn}} = 0.708$  ( $\sigma = 0.455$ ), Pos. 3  $\bar{x}_{1\text{-CatCor Syn}} = 0.68$  ( $\sigma = 0.467$ ), Pos. 4  $\bar{x}_{1\text{-CatCor Syn}} = 0.081$  ( $\sigma = 0.273$ ), EOS-Marker  $\bar{x}_{1\text{-CatCor Syn}} = 0.116$  ( $\sigma = 0.32$ ).

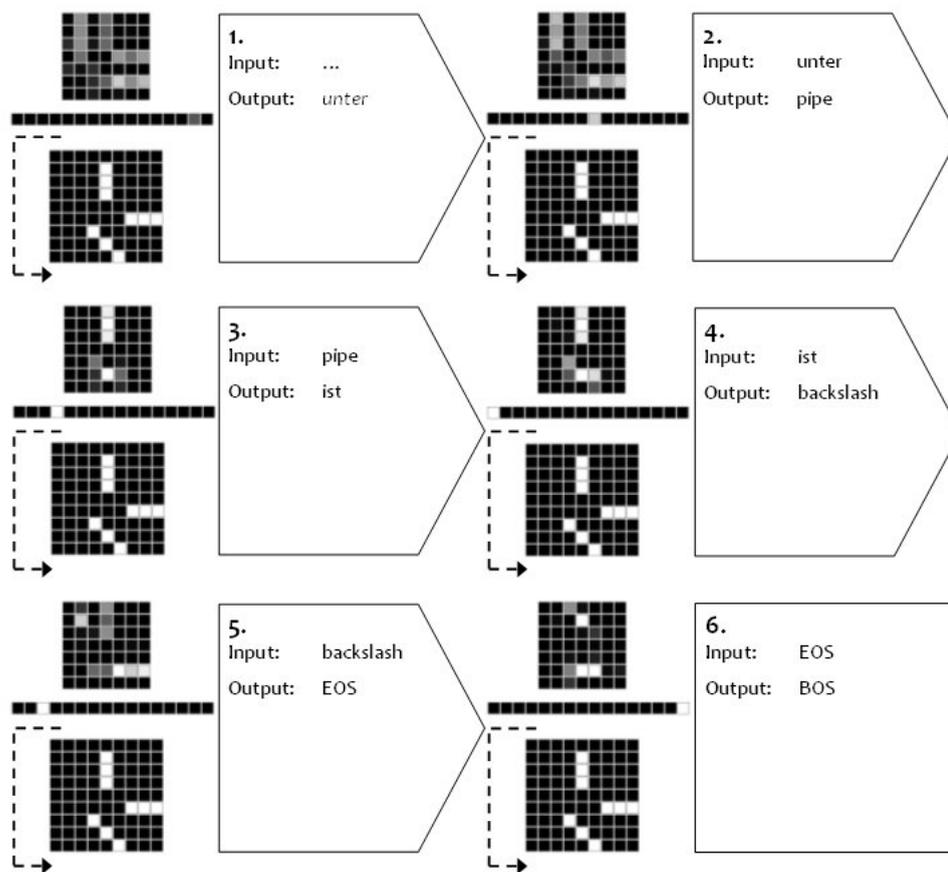


Abbildung 10.13.: SynSem I: Beispiel zur beschreibenden Produktion eines Satzes mit Feedback-Erweiterung der Architektur.

darüber hinwegtäuschen, dass die Performanz für diesen Test-Modus insgesamt betrachtet nur mäßig erfolgreich war. Das gegebene Beispiel ist also eher eine Ausnahme. Die Pfeile in der Abbildung erinnern daran, dass der netzwerkinterne Input über die Feedback-Schleife online während der Verarbeitung generiert wird: Immer die maximal aktivierte Unit wird als neuer Input eingeschleift. Im Beispiel sehen wir eine Situation mit drei Input-Objekten in den Relationen PIPE ÜBER BACKSLASH, PIPE LINKSÜBER MINUS und MINUS RECHTSÜBER BACKSLASH. Untersuchen wir hierfür die Entwicklung des Outputs der visuellen und der sprachlichen Route:

1. Als erster Output der sprachlichen Route entwickelt sich das Adverb *unter*, parallel dazu zeigen sich im visuellen Output die Objekte PIPE in den Relationen ÜBER und LINKSÜBER sowie MINUS in den Relationen RECHTSUNTER und RECHTSVON. Bis auf die letzte Relation sind dies alles valide Aspekte der Situation.
2. Die Andeutungen im visuellen Output bleiben auch auf Position 2 erhalten, es zeigt sich aber keine Reaktion auf den Input *unter*. Diese zeigt sich dafür im sprachlichen Output mit der Aktivierung des Nomens *pipe*, das in der gegebenen Situation als einziges das korrekte Objekt einer entsprechenden ÜBER/UNTER-Relation benennt – eine sprachlich wie situational-semantisch korrekte Vorhersage.
3. Auf den Input *pipe* folgt korrekt die Vorhersage *ist*. Nun folgt auch verspätet die Reaktion des visuellen Outputs auf die UNTER-Relation: Es deutet sich die Aktivierung von BACKSLASH in entsprechender Positionierung an. Falsche Aktivierungen anderer Positionierungen sind weitgehend reduziert, zudem wird PIPE nun eindeutig in der ÜBER-Relation abgebildet.
4. Nun folgt semantisch und syntaktisch korrekt die Vorhersage von *backslash*. Im visuellen Input verstärkt sich die Aktivierung von BACKSLASH in der UNTER-Relation, allerdings in Konkurrenz zum MINUS mit derselben Positionierung.
5. Es wird korrekt der *EOS* vorhergesagt. Allerdings verschlechtert sich die korrespondierende Abbildung im visuellen Output. Sowohl BACKSLASH als auch die UNTER-Relation gehen verloren.
6. Abschließend folgt die korrekte Vorhersage von *BOS* mit nochmals veränderter und weiterhin inkorrekt abgebildeter Abbildung im visuellen Output.

Das Beispiel zeigt, dass mitunter syntaktisch korrekte sowie semantisch kohärente Wortfolgen produziert werden, die zu dem zu beschreibenden visuellen Kontext korrespondieren – in diesem Fall der Satz *BOS unter pipe ist backslash EOS*. Dabei wirkt sich der visuelle Input sowohl auf die visuelle als auch die sprachliche Verarbeitung aus. Allerdings zeigt sich im visuellen Output, dass sich das Fehlen des expliziten sprachlichen Inputs negativ auf die Konsistenz des abgebildeten Situationsmodells auswirkt – auch wenn es scheinbar nur das satzinitiale *BOS* ist, das tatsächlich fehlt. Hierzu ist anzumerken, dass die eindeutigen Aktivierungen bei der Wortvorhersage ein Resultat der Winner-takes-all-Funktion des Output-Layers der sprachlichen Route sind. Welche anderen Aktivierungen an diesen Layer propagiert werden, wird dadurch nicht mehr ersichtlich. Die eher mäßige Performanz für den visuellen Output deutet jedenfalls darauf hin, dass Sprache in diesem Modell eine wichtige stützende Funktion für die Etablierung einer eindeutigen Interpretation der gegebenen Situation innehat.

**Detailanalyse: Sprachliche Route** Für die Analyse der Performanz bei der Wortklassen-Vorhersage – als Realisierung der Wort-Produktion – haben wir auf den Satzpositionen die Aktivierung der Wortklassen beider Satztypen (ORO und ROO) als Fortführung geprüft, da hinsichtlich des Satztyps bei der freien Produktion keine Vorgaben gemacht wurden. Dazu haben wir im Sinne des Fehlers 1-CatCor Syn die Wortklasse des maximal aktivierten Worts mit den auf dieser Position möglichen Wortklassen verglichen. Das Diagramm 10.14 zeigt anhand des Fehlerverlaufs bei zwei oder drei Input-Objekten<sup>22</sup> recht deutlich, dass die Leistung bei der Produktion sehr mäßig ist. Niedrige Fehlerwerte finden sich nur auf den ersten beiden und den letzten beiden Positionen, die Vorhersagen auf den inhaltlich sehr relevanten Positionen 3 und 4 hingegen sind wenig erfolgreich. Wir haben die produzierten Strukturen näher überprüft. Neben wenigen korrekten Strukturen werden am häufigsten die folgenden beiden falschen Strukturen gebildet: (a) Nomen Nomen Verb Nomen EOS BOS, (b) Adverb Nomen EOS BOS EOS BOS. Die erste Struktur entspricht dem Satztyp ROO mit falscher Eröffnung, die ersten vier Positionen der zweiten Struktur entsprechen dem Abschluss des Satztyps ORO. Zumindest partiell werden hierbei also korrekte Wort-Sequenzen gebildet. Als weiterer Fehler erscheint mitunter ein EOS auf der initialen Position; dies deutet darauf hin, dass gerade die fehlende Satzeröffnung für das Netzwerk ein Problem darstellt. Inwiefern wird mit den maximal aktivierten Wörtern auf Objekte und Relationen verwiesen, die in den Situationen erscheinen? Wir haben zur Beantwortung dieser Frage die produzierten Wörter (nur Nomen und Adverbien) mit den möglichen situationalen Referenzen verglichen. Bei zwei Input-Objekten korrespondieren über 85% der produzierten Wörter mit den möglichen situationalen Referenzen.<sup>23</sup> Insofern hat zumindest die lexikalische Auswahl bei der Produktion beschreibenden Charakter.

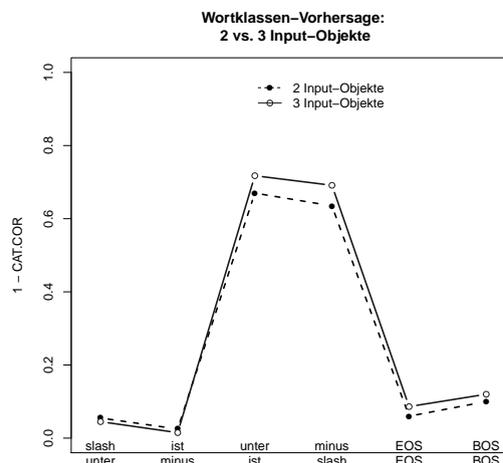


Abbildung 10.14.: SynSem I: Ergebnisse für die Vorhersage-Performanz für die Wortklasse bei beschreibender Satzproduktion (1-CatCor Syn).

<sup>22</sup>Wir können in den folgenden Diagrammen keine Gruppierung nach Satztypen (ORO und ROO) vornehmen, da der Output willkürlich die eine oder andere Satzform adaptiert oder aber durch Strukturfehler keinem der beiden Satztypen eindeutig zugeordnet werden kann.

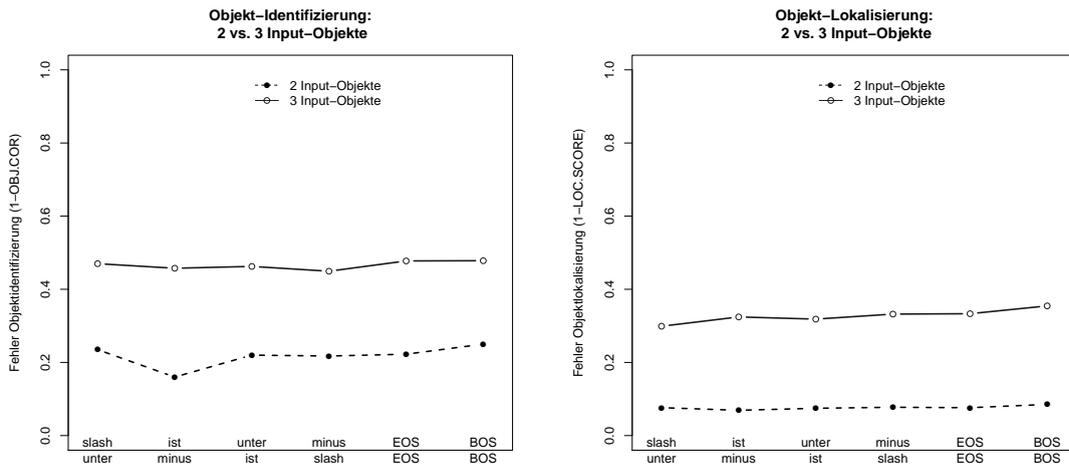
<sup>23</sup>Im Falle von drei Input-Objekten ist die Korrespondenz noch höher, allerdings sind hier zumeist fast alle Relationen und Objekte möglich.

**Detailanalyse: Visuelle Route** Neben der Produktion im sprachlichen Output haben wir die Performanz der visuellen Route analysiert (vgl. Diagramme in Abb. 10.15). Werden vorwiegend korrespondierende Wörter aktiviert und zumindest partiell korrekte Strukturen produziert, so sollte sich gleichermaßen ein zur beschriebenen Situation korrespondierendes Situationsmodell etablieren. Die Fehlerverläufe sind in dieser Beziehung allerdings eher ernüchternd: Objekt-Identifizierung und -Lokalisierung des erzeugten Situationsmodells erfolgen scheinbar gänzlich unabhängig von der Satzproduktion – zumindest zeigen sich keine Einflüsse der verschiedenen Satz-Positionen auf die Fehlerwerte. Insgesamt sind die Werte für die Bedingung mit zwei Objekten etwas besser, was daran liegt, dass die Abbildung der Situation prinzipiell unabhängig von der sprachlichen Konkretisierung erfolgen kann. Beim Fehler für die Erfassung möglicher Relationen (PosRel Sem) zeigt sich im Verlauf des Satzes sogar eine sukzessive Verschlechterung für den visuellen Output. Diese Ergebnisse decken sich recht gut mit den Verhältnissen in Beispiel 10.13.

**Zusammenfassung** Die Ergebnisse betreffend Sprachproduktion zeigen, dass die Generierung syntaktischer Sequenzen in unserem Modell alles andere als ausgereift ist. Zwar werden weitgehend diejenigen Wörter verwendet, die auf Aspekte referieren, die in der zu beschreibenden Situation zugegen sind. Und die selteneren tatsächlich korrekten Sequenzen beschreiben tatsächlich den Sachverhalt des situationalen Kontexts. Zudem werden größtenteils partiell korrekte Sequenzen produziert. Allerdings lassen sich aus diesen nur bedingt Beschreibungen der jeweiligen Situation herauslesen, denn selbst wenn die lexikalisch korrekten Elemente verwendet werden, werden dabei keine vollständigen, kohärenten und zur gezeigten Relation korrespondierende Aussagen gebildet. Der sprachliche Input scheint in unserem Modell eine wichtige Stützfunktion für die korrekte sprachliche und visuelle Verarbeitung zu erfüllen. Die partiell korrekten Sequenzen deuten darauf hin, dass beim Fehlen sprachlichen Inputs die Orientierung über die gesamte Sequenz hinweg fehlt – das Gedächtnis des Systems für Sequenzen scheint also stark beschränkt zu sein, das Netzwerk weiß sozusagen nicht, wo es sich innerhalb der Sequenz befindet. Dabei wirkt sich vor allem das Fehlen eines Startsignals für die Eröffnung der Sequenzen negativ auf die Performanz aus. Für die Erfüllung einer hinreichend erfolgreichen Produktionsleistung scheint jedenfalls irgendein zusätzlicher strukturbildender Mechanismus benötigt zu werden, der die Entwicklung der Sequenz überwacht und über den Entscheidungen zur konkreten sprachlichen Realisierung einer Beschreibung getroffen werden.

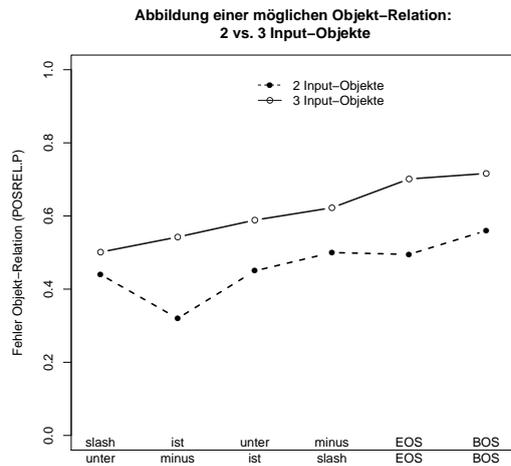
### 10.2.4. Diskussion von SynSem I

Fassen wir die Ergebnisse zusammen: Das Modell SynSem I leistet die Integration sequenzieller sprachlicher und statischer visueller Aufgaben und Informationen. Dadurch können sich die Informationen der beiden Routen gegenseitig als Stimulus und Beschränkung dienen. Dies führt zum einen dazu, dass die Performanz für eine Aufgabe modalitätsübergreifend durch die jeweils andere Route angesteuert werden kann. Diese Performanz kann im einen Fall als (entkoppeltes oder kontextfreies) Sprachverstehen und im anderen als (situational referierende) Sprachproduktion beschrieben werden. Zum anderen führt die gegenseitige Beschränkung dazu, dass die zusätzliche situationale oder sprachliche Information für Inferenzen genutzt werden können. Im Falle der Wortvorhersage als Antizipation



(a) Objekt-Identifizierung (1-ObjCor Sem)

(b) Objekt-Lokalisierung (LocScore Sem)



(c) Mögliche Relation (PosRel Sem)

Abbildung 10.15.: SynSem I: Performanz der visuellen Route bei der beschreibenden Satz-Produktion: Objekt-Identifizierung (1-ObjCor), Objekt-Lokalisierung (LocScore) und Fehler der Abbildung korrekter Objekt-Relationen (PosRel). Vergleich der Ausgangssituationen mit zwei vs. drei Input-Objekten.

oder modalitätsübergreifendes Priming, im Falle der visuellen Situationskonstruktion als kontextuelle Inferenz oder visuelle Aufmerksamkeitslenkung. Diese gegenseitige Nutzung bei der situierten Sprachverarbeitung zeigt sich besonders deutlich im Vergleich zum kontextfreien Verstehen. Hier erfolgt die Konstruktion der visuellen Situation inkrementell als sukzessiver Aufbau, bei dem jedes Inhaltswort zu einer Konkretisierung führt; die Wortvorhersage erfolgt als rein auf Kollokation basierende Wortklassen-Vorhersage. Bei der situierten Verarbeitung hingegen ergibt sich bei der Wortvorhersage eine referenziell-semanticke Restriktion, die schon eher tatsächlicher Interpretation entspricht. Und bei der Situationskonstruktion zeigen sich schlagartige Konkretisierungen der Repräsentation durch die Nutzung der sich gegenseitig ergänzenden Informationen. Zudem erfüllt der sprachliche Input in diesem Modell eine stark stützende Funktion für die Verarbeitung, die sich besonders bei seinem Ausbleiben im Modus der Sprachproduktion zeigt.

### 10.2.4.1. Theoretische und empirische Bezüge

Die Ergebnisse der situierten Verarbeitung stimmen mit empirischen Beobachtungen zur Sprachverarbeitung überein. Unseren Simulationen zufolge werden bei der unmittelbar-inkrementellen Anpassung des Situationsmodells alle Informationen gleichermaßen berücksichtigt – neben dem sprachlichen Input werden gleichermaßen auch Informationen bzw. Beschränkungen des visuellen Kontexts unmittelbar integriert. Dies entspricht den empirischen Beobachtungen, die in verschiedenen Visual-World-Studien gemacht wurden (z.B. Tanenhaus et al., 1995; Altmann & Kamide, 1999; Kamide, Scheepers & Altmann, 2003; Altmann & Kamide, 2007), und beispielsweise explizit im Coordinated Interplay Account thematisiert werden (Knoeferle & Crocker, 2006). Die Leistungsfähigkeit des Modells – und das spezifische Verhalten mit weitergehenden Leistungen wie Inferenzen, Aufmerksamkeitslenkung und modalitätsübergreifendem Priming – entwickelt sich dabei nicht durch extern vorgegebene oder explizit implementierte Mechanismen, sondern als emergentes Produkt des situierten Spracherwerbs durch die Aufgaben-Koordination und -Integration und die prototypisierte Erfassung der visuellen Situationen. Damit stellt das Modell eine Alternative zum Ansatz im Modell CIA-Net (Mayberry et al., 2009) dar und entspricht den Annahmen von Altmann und Mirkovic, die Antizipation und Lenkung visueller Aufmerksamkeit als Produkt eines übergreifenden repräsentationalen Substrats sehen (Altmann & Mirkovic, 2009, S. 18):

This account does not require an external mechanism [...], because attention is instantiated within the same representational substrate as linguistic and nonlinguistic information [...]; in other words, different states of this representational substrate represent the attentional modulation that drives eye movements.

Vor dem Hintergrund der Modellierung von Embodied Language Processing und von situiertem Spracherwerb können mehrere Parallelen gezogen werden. Zunächst einmal können wir Sprachverstehen in unserem Modell im Sinne von Zwaan (2004) oder Barsalou (1999) als ‚mentale‘ Simulation (multi-)modaler Erfahrungsderivate beschreiben. Anhand linguistischer Indizierung erfolgt im Integrations-Layer ein Re-Enactment von Korrelaten früherer Wahrnehmungserfahrungen. Umgekehrt kann Sprachproduktion in unserem Modell als Re-Enactment sprachlicher Erfahrungsderivate durch nichtsprachliche Informationen beschrieben werden. Durch seine Rekurrenz steuert der Integrations-Layer die

Konstruktion der Simulation von ‚mentalen‘ Situationsmodellen, die zur sprachlichen Beschreibung korrespondieren. Durch die hinzukommenden Informationen während des Satzverstehens wird dieses ständig aktualisiert und dabei konkretisiert, revidiert oder erweitert. Durch unterspezifizierte Repräsentationen kann im Sinne inkrementeller Verarbeitung auch bei unvollständiger oder vager Information ein konsistentes Situationsmodell erzeugt werden. Der Umfang der Änderungsprozesse beim inkrementellen Aufbau des Situationsmodells könnte eventuell als Ansatz für die Bemessung von Komplexität beim Sprachverstehen herangezogen werden. Einen ähnlichen Bezug schlägt zum Beispiel Rohde (2002) mit der Bemessung der Änderung des Message-Gestalt-Layers als Indikator von Komplexität im CSCP-Modell vor.

Mit dem Modell wird auch der situierte Spracherwerb simuliert. Genähert an die Annahmen von Glenberg und Robertson (2000) lassen sich die Vorgänge im Modell so beschreiben, dass die schematisierten bzw. prototypisierten Korrelate aus der Kategorisierung visueller Situationen mit linguistischen Formen indiziert werden. Dies erfolgt nicht auf holistischen Repräsentationen der Situationen, vielmehr sind einzelne Aspekte der Situationen (Objekte, Lokalisierungen) Gegenstand der Indizierung. Die Abbildung von Form auf Bedeutung etabliert sich in unserem Modell an der Schnittstelle – dem Integrations-Layer. In diesem sowie in seinen Gewichtsstrukturen sind das konzeptuelle Wissen des Modells gespeichert und die sprachliche Bedeutung verankert. Durch ihre Integration mit den Derivaten der Situations-Kategorisierung ist die Sprache ge-groundet, sie wird bedeutungshaft, weil sie korrespondierendes situationales Wissen ansteuern kann. Die Etablierung von Eins-zu-eins-Korrespondenzen zwischen einzelnen lexikalischen Elementen des Satzes und den verschiedenen Aspekten bzw. Bestandteilen der korrespondierenden Situation lässt sich am ehesten als selbstorganisierendes situationsübergreifendes statistisches Lernen (Gleitman, 1990; Davis, 2002) beschreiben. Unser Modell bietet damit eine selbstorganisierende konnektionistische Alternative zu den bisherigen algorithmischen und hybriden Ansätzen (z.B. Siskind, 1996; Roy, 1999; Yu & Ballard, 2003, 2004; Yu, 2006).

#### 10.2.4.2. Probleme und Einschränkungen

Trotz aller positiven Ergebnisse gibt es im Modell SynSem I eine zentrale Schattenseite des Ansatzes: Viele der Eigenschaften beruhen auf der Prototypisierung der Input-Szenen – und gerade diese Prototypisierung haben wir explizit in den Target-Vektoren der visuellen Route enkodiert.<sup>24</sup> Das ist dem Vorgehen ganz ähnlich, das im Modell CIA-Net (Mayberry et al., 2009) umgesetzt wurde – in diesem werden zwei Input-Propositionen auf eine ausgewählte Ziel-Proposition abgebildet. Dadurch, dass unsere Prototypen handkodiert vorgegeben sind, können wir nicht entscheiden, welche Performanzaspekte letztlich von diesen Vorgaben abhängen bzw. ob das Modell auch gänzlich selbstorganisiert eine

<sup>24</sup>Genau genommen ist zudem die Anzahl möglicher zu einem Satz korrespondierender Ziel-Vektoren durch die Idealisierung und Prototypisierung der Situationen wesentlich geringer als die Anzahl der Input-Situationen: Es ergeben sich kombinatorisch nur 52 mögliche Output-Repräsentationen – das Netzwerk muss also nur eine sehr beschränkte Abbildungsleistung erbringen. Allerdings zeigt das Modell auch, dass es sehr wohl die Informationen des visuellen Inputs kategorisiert hat und für die Verarbeitung und Antizipation von Wortvorhersage und Muster-Assoziation nutzen kann. Es leistet also wesentlich mehr als nur die Abbildung auf die 52 möglichen Target-Repräsentationen.

entsprechende Prototypisierung leisten könnte, die die gezeigten Leistungen ermöglicht. Wir haben unserem Modell diesen wichtigen Verarbeitungsschritt abgenommen, ohne eine konkrete Rechtfertigung sowohl für die Erleichterung als auch für die Gestaltung der Prototypisierung zu nennen. Folglich wird es für die Weiterentwicklung des Ansatzes wichtig, entweder schlüssige Begründungen für diese Entscheidungen zu liefern oder aber zu zeigen, dass Netzwerke auch selbstorganisierend gleichwertige Prototypisierungen leisten können. Um das zweitgenannte Ziel zu verfolgen, müssen wir entsprechende Leistungsmerkmale in Modellen mit unüberwachten oder selbstüberwachten Lernverfahren nachweisen und in unser Modell aufnehmen.

Abgesehen hiervon bleibt festzuhalten, dass die Komplexität der situierten Sprachverarbeitungsaufgabe in unseren Simulationen in mehrfacher Hinsicht sehr beschränkt bleibt und sich die Beschränkungen nur teilweise rechtfertigen lassen:

- Die visuellen Szenen sind sehr reduziert und primitiv: Die Objekte besitzen keine Eigenschaften, über die sie gemeinsam kategorisiert und kategorial voneinander unterschieden werden könnten. Die Situationen stellen keine Bewegung und keine Interaktion von Objekten dar, zudem bestehen keine Temporalität und Kausalität und kein sonstiger Zusammenhang zwischen situationalen Zuständen.
- Das Modell bleibt auf die Domänen der sprachlichen und der visuellen Verarbeitung beschränkt, andere Modalitäten werden komplett ausgeblendet. Allerdings nehmen wir an, dass hinzukommende Modalitäten weniger ein Hindernis für die Sprachverarbeitung, sondern unserem Ansatz folgend vielmehr eine Unterstützung der Erwerbs- und Verarbeitungsprozesse durch zusätzliche Informationen und Constraints darstellen würden.
- Die sprachlichen Äußerungen beziehen sich bei vorhandenem visuellem Input immer auf die korrespondierende Szene. Allerdings ist dies bei kindgerichteter Sprache durchaus ebenfalls gegeben, denn sie ist wesentlich situationsbezogener und referentieller (Verwendung vieler Nomina sowie unterstützender deiktischer Gesten) als Erwachsenensprache und geht in der Regel einher mit der Etablierung einer gemeinsamen gerichteten Aufmerksamkeit.
- Die verwendeten Sätze weisen eine stark beschränkte Anzahl sehr einfacher Strukturen auf, die nur eine Länge von vier Inhaltswörtern umfassen. Aber auch diese Vereinfachung kann man mit Blick auf die Merkmale kindgerichteter Sprache rechtfertigen.

### 10.3. Modellierung SynSem II

Der Modellierungsansatz SynSem I hat interessante Ergebnisse erbracht, die darauf hindeuten, dass spezielle Verhaltensweisen des Sprachverstehens emergent aus der Integration von sprachlicher Form und situationaler Bedeutung entstehen könnten. Allerdings besteht die Möglichkeit, dass sich diese Ergebnisse direkt auf unsere Manipulationen der Zielstrukturen der visuellen Route zurückführen lassen. Um diese Möglichkeit auszuschließen, benötigen wir eine Version von SynSem, in der das Wissen vollständig selbstüberwacht erworben wird. Unser Ausgangspunkt hierfür ist, dass die visuelle Route ausschließlich

rein auto-assoziative Transformationen leisten und folglich die Input-Situation auf eine identische Target-Repräsentation abbilden muss. Unter Verwendung der Aufgaben Wortvorhersage und situationale Auto-Assoziation erhält das Modell keine Informationen mehr, die nicht auch in seiner Umwelt gegeben sind. Das Projekt SynSem II ist ein erster Versuch, zu zeigen, dass ein solches Modell das gleiche Verhalten und die gleiche Leistungsfähigkeit oder zumindest eine vergleichbare Performanz aufweist, wie sie für das gestützte Modell im Projekt SynSem I gezeigt werden konnte. Wir haben für SynSem II verschiedene alternative Architekturvarianten getestet und Anpassungen der Stimuli vorgenommen. Dabei haben wir auch verschiedene Erweiterungen vorgenommen, die unter anderem eine bessere Untersuchung der Systematizität des Modells erlauben. Wir beginnen mit der Erläuterung der Stimuli und gehen dann zu den vorgenommenen Variationen der Architektur über. Nach der Darlegung der Trainings- und Test-Stimuli und der Simulations-Details präsentieren wir die Ergebnisse verschiedener Simulations-Studien im Projekt SynSem II.

### 10.3.1. Stimuli und neue Architektur

Für die Gestaltung einer auto-assoziativen Variante des Projekts müssen wir einige Aspekte der Stimuli verändern, einerseits, um der Vorgabe nicht vorgegebener Prototypen gerecht zu werden, andererseits, um die Aufgabe überhaupt bewältigbar zu gestalten. Der Input der visuellen Route ist im zweiten Projekt prinzipiell gleich aufgebaut wie in SynSem I: Wir verwenden jeweils eines, zwei oder drei Objekte aus der Auswahl der vier bekannten Objekte MINUS, PIPE, SLASH und BACKSLASH. Diese werden frei auf dem Input-Gitter verteilt und bilden dabei eine räumliche Relation zueinander. Die Target-Aktivierung am Output-Layer der visuellen Route bildet exakt die gleiche Relation ab, die am Input vorliegt (vgl. Abb. 10.16). Damit werden strukturelle Informationen durch die Transformation des Inputs auf einen vorstrukturierten Output ausgeschlossen. Stattdessen verfügt das Material aber über eine implizite Informationsstruktur: Die Objekte erscheinen mit verschiedener Wahrscheinlichkeit auf dem Unit-Gitter und spiegeln damit eine probabilistische Prototypizität der Objekt-Relation wider. Dies entspricht dem Vorgehen in früheren Modellen, in denen die Erfassung von Prototypizitäts-Effekten simuliert wird (vgl. z.B. McClelland & Rumelhart, 1985; Plunkett et al., 1992). Zu diesem Zweck haben wir den Retina-Layer stark erweitert und eine Einteilung in prototypische Felder vorgenommen, die in Abbildung 10.17 veranschaulicht wird. Auf Grundlage dieser Feldeinteilung haben wir verschieden komplexe Stimuli erzeugt, die den Grad der Abweichung der Objekte von der Ideal-Position und der Positionierungsfreiheit der Objekte bei der Herstellung von Objekt-Relationen variieren.<sup>25</sup>

#### 10.3.1.1. Eigenschaften der Stimuli

Die Retina ist in insgesamt 9 Felder aufgeteilt, die jeweils 7x7 Units umfassen. Die Felder überlappen an ihren Grenzen um jeweils zwei Units. Damit umfasst die Retina insgesamt 17x17 Units (insgesamt 289 Units). Die 9 Felder besitzen einen Mittelpunkt (in Abb. 10.17

<sup>25</sup>Das CommonLisp-Tool zur Generierung der probabilistisch kontrollierten Stimuli wurde von Benjamin Udiljak im Rahmen seiner Diplomarbeit entwickelt (Udiljak, 2010). Dies ermöglichte die Umsetzung dieser Folgestudie – herzlichen Dank hierfür an dieser Stelle.

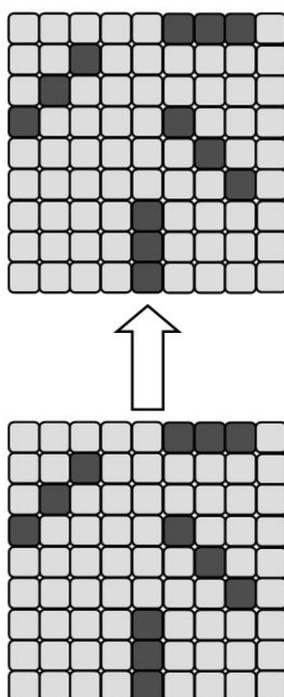


Abbildung 10.16.: Visuelle Stimuli der Mikrowelt für die Simulation SynSem II. Aufgabe der Transformation ist die Auto-Assoziation der Situation mit beliebig angeordneten Objekten auf eine identische Repräsentation.

durch die Position der Feldzahl bestimmt). Dieser Mittelpunkt bestimmt die prototypische Positionierung der Objekte. Die Prototypizität der dargestellten Relationen wird durch die Belegung der Felder bestimmt: Bei prototypischer Relation liegen sich die relevanten Objekte genau gegenüber. So müssen sich etwa zwei Objekte für die Relation LINKSÜBER in den Feldern 1 und 9 befinden und für die Relation RECHTSVON in den Feldern 4 und 6. In den verschiedenen komplexen Varianten der Stimuli-Materialien weichen die Objekte und Relationen in verschiedenem Maße von diesen idealen Positionen ab. Wir unterscheiden zwei verschiedene Formen von Abweichung: Positionsrauschen und Relationsrauschen. Als Positionsrauschen bezeichnen wir die Verschiebung von Objekten innerhalb eines Felds in beliebiger Richtung weg vom Zentrum des Felds; bei einer Feldgröße von  $7 \times 7$  Units ist eine Verschiebung von bis zu zwei Units möglich. Jedes Feld kann dabei nur durch ein Objekt belegt werden. Dadurch nimmt ein drittes Objekt automatisch ein Feld ein, das nicht der idealen gegenüberliegenden Positionierung einer prototypischen Relation entspricht. Als Relationsrauschen bezeichnen wir die Verschiebung von Objekten auf andere Felder, mit denen nicht mehr die ideale gegenüberliegende Positionierung erfüllt wird. Liegt Objekt A auf Feld 1, so kann Objekt B für die Relation RECHTS UNTER auf den Feldern 5, 6, 8 oder 9 liegen. Zudem ist auch dann die Relation RECHTS UNTER möglich, wenn Objekt A auf Feld 2 (und Objekt B entsprechend auf Feld 6 oder 9), auf Feld 4 (und Objekt B auf Feld 8 oder 9) oder auf Feld 5 (und Objekt B entsprechend auf Feld 9) liegt. Positions- und Relationsrauschen können miteinander auftreten, ein Objekt kann also auf einem nicht-idealen Feld liegen und dort zudem nicht-ideal positioniert sein. Zudem kann durch die Verschiebung mehr als ein Objekt in einem Feld liegen.

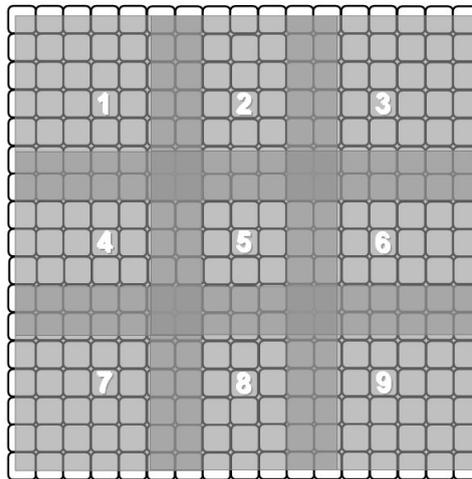


Abbildung 10.17.: Retinale Felder für die Simulation SynSem II. Aufteilung in Felder prototypischer Orientierung mit Kern-, Rand- und Überschneidungsbereichen. Diese prototypische Orientierung betrifft nur die darauf bezogene sprachliche Beschreibung.

In der primitivsten Variante verwenden wir nur Objekte auf Ideal-Positionen (jeweils im Zentrum der dargestellten Felder), zudem erscheinen die relevanten Objekte einer beschriebenen Relation immer in den idealen gegenüberliegenden Feldern. So müssen sich etwa zwei Objekte für die Relation LINKSÜBER in den Feldern 1 und 9 befinden und für die Relation RECHTSVON in den Feldern 4 und 6. Diese Variante ist trivial, da sie schlichtweg die Stimuli so weit vereinfacht, dass keine speziellen Strukturvorgaben mehr nötig sind. Deswegen wird diese Version nicht weiter behandelt. Die komplexere Variante A verwendet die gleiche Vorgabe der Positionierung von jeweils zwei Objekten in gegenüberliegenden Feldern, führt hierfür aber Positionsrauschen ein. Die dargebotenen Objekte weichen in der im Folgenden dargestellten Variante in verschiedenen Graden von der Ideal-Position ab: Eine Abweichung um eine Unit in beliebiger Richtung findet in 15% der Fälle statt, eine Abweichung um zwei Units nur noch in 5% der Fälle.<sup>26</sup> In der komplexesten Variante B haben wir zusätzlich zum Positionsrauschen in verschiedenem Umfang Relationsrauschen eingeführt. In drei Untervarianten haben wir den Einfluss von 10%-, 30%- und 50%-Abweichungen auf die Performanz untersucht.

### 10.3.1.2. Merkmale der neuen Architektur

Die Herausforderung für die neue Architektur ist, durch Reduzierung und Idealisierung von Objekten und Relationen eine Prototypisierung der Input-Situationen herzustellen. Diese Prototypisierung liegt vielen Leistungen von SynSem I zugrunde, war hier aber durch vorgegebene Target-Vektoren gegeben. In SynSem II soll sie selbstorganisierend und ohne von außen vorgegebene Spezifikationen vom Netzwerk selbst geleistet werden. Vorbilder hierfür sind etwa die Modelle von McClelland und Rumelhart (1985) oder Plunkett et al.

<sup>26</sup>Dies sind keine starken Anteile abweichender Fälle, genügen aber, um unsere Fragestellungen zu untersuchen; eine größere Anzahl von Abweichungen wäre geeignet, um die Robustheit und die Grenzen der Generalisierungsfähigkeit des Netzwerks zu untersuchen.

(1992). Allerdings sind die Anforderungen an diese Modelle hinsichtlich der Voraussetzungen wesentlich geringer, da die Stimuli in verschiedenen Aspekten einfacher waren – vor allem in ihrer Enkodierung von Raum und Identität. Eine wesentliche Hürde, die die meisten unserer Versuche vor Probleme gestellt hat, war die Herstellung von Positions-Invarianz – Objekt A bleibt Objekt A, egal auf welcher Position es erscheint, und Relation X bleibt Relation X, egal auf welchen Positionen die verglichenen Objekte erscheinen. Modellierungen mit der ursprünglichen Architektur haben sich hierfür als ungeeignet erwiesen. Wie erläutert, sind hier räumlich versetzte Aktivierungsmuster maximal unähnlich, da keine räumliche Beziehung zwischen den Units herrscht – hinsichtlich des Verknüpfungsschemas besteht keinerlei Beziehung zwischen ihnen. Die Ähnlichkeiten der Objekte und Relationen werden durch die jeweiligen gemeinsamen Target-Repräsentationen hergestellt. Dies ist mit den neuen Stimuli und der auto-assoziativen Transformationsaufgabe nicht mehr gewährleistet.

Wir haben zunächst verschiedene Erweiterungen der Ausgangs-Architektur getestet. In all diesen Erweiterungen der ersten Versuchsphase ging es darum, auto-assoziative Verknüpfungsmuster zu integrieren, durch die nach dem Vorbild des „Cats, Dogs and Bagels“-Modells (McClelland & Rumelhart, 1985) aus den atypisch verschobenen Input-Situationen gemeinsame prototypische Objekt-Positionierungen und -Relationen abgeleitet werden sollten. All diese Versuche erbrachten nicht die gewünschten Ergebnisse (vgl. Udiljak, 2010), aus der Performanz konnte mitunter nicht einmal auf eine Lösung grundlegender Funktionen geschlossen werden. Folglich verzichteten wir auf eine nähere Darstellung dieser Versuche. Der Grund für die unzureichende Performanz ergibt sich daraus, dass die Verzerrungen der Prototypen in den Vorbild-Modellen nur einen Teil der jeweils zugrundeliegenden Repräsentation betraf, während der größte Teil der Repräsentation aufrechterhalten wurde. Alle zu einem Prototyp gehörenden Repräsentationen behielten also eine große Schnittmenge zu ihrem Prototyp. In unserem Modell hingegen können die Repräsentationen wie erläutert maximal unterschiedlich sein und keinerlei Schnittmenge zum Prototypen erhalten, während sie mitunter wesentlich größere Ähnlichkeit zu konkurrierenden Prototypen aufweisen. Unter dieser Voraussetzung hatten die auto-assoziativen Erweiterungen unseres Modells keine Auswirkung auf die Lösung der Aufgabe.

Etwas erfolgreicher war die folgende Variante: Durch die vollständigen Verknüpfungen und den rekurrenten Integrations-Layer sollte unsere Architektur eigentlich alle Voraussetzungen mitbringen, um als auto-assoziatives Netzwerk zu funktionieren. Um dieses Potenzial für die gestellte Aufgabe nutzen zu können, haben wir an einem anderen Aspekt des Modells angesetzt: der Herstellung räumlicher Ähnlichkeit während der frühen visuellen Verarbeitung durch die Erzeugung von Sensitivität für Nachbarschaftsbeziehungen zwischen Input-Mustern. Hierzu haben wir den Input-Layer der visuellen Route um die Funktion rezeptiver Felder erweitert. Diese Erweiterung soll die Erzeugung positionsinvarianter Objekt-Konstanz und prototypisierter räumlicher Relationen unterstützen. Außer dem hinzugefügten rezeptiven Layer bleibt die Architektur aber identisch zur Version, die wir für das Projekt SynSem I dargestellt haben. Im neuen Netzwerk erfolgt die Propagierung von Aktivierung in der visuellen Route also vom visuellen Input-Layer zum rezeptiven Layer und von dort weiter in den ersten Kompressions-Layer. Die übrigen Aktivierungs- und Fehler-Propagierungen bleiben wie gehabt.

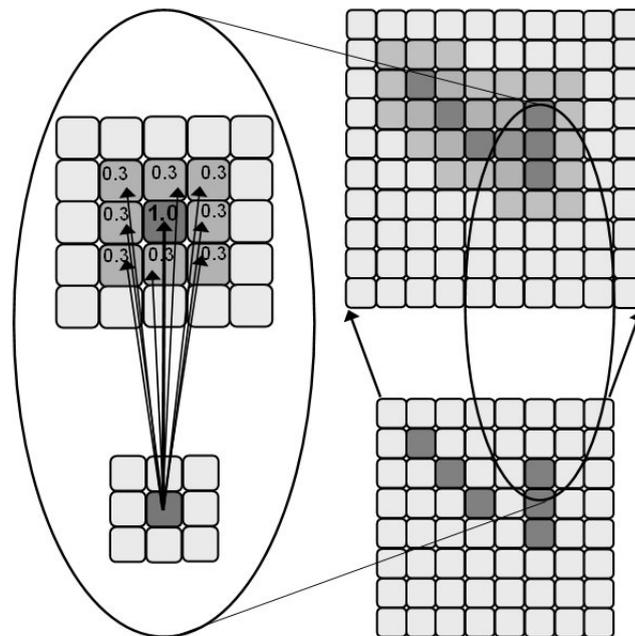


Abbildung 10.18.: Einführung rezeptiver Felder in die Architektur zur Vorverarbeitung des visuellen Inputs und zur Herstellung räumlicher Nachbarschaft zwischen den zweidimensional angeordneten Input-Units.

Rezeptive Felder sind eine Möglichkeit, in einem räumlich organisierten (z.B. zweidimensionalen) Layer die Nachbarschaft zwischen Units herzustellen und dadurch räumlich dargestellte Objekte, Relationen und Ereignisse besser zu verarbeiten.<sup>27</sup> Die rezeptiven Felder für die SynSem-Architektur wurden nach dem Vorbild des Modells von Plunkett et al. (1992) implementiert. Die Verknüpfung des retinalen Inputs zum Layer für die räumlich-visuelle Verarbeitung erfolgt selektiv über fixierte Gewichte (vgl. Abb. 10.18): Die Units des Input-Layers sind mit den deckungsgleichen visuellen Units über ein Gewicht der Verbindungsstärke 1.0 verbunden; die diese Unit direkt umgebenden Units werden mit einem Gewicht der Verbindungsstärke 0.3 versehen. Somit ist jede Input-Unit mit genau 9 visuellen Units verknüpft, die dessen rezeptives Feld bilden. Unter diesen Voraussetzungen approximieren wir ein rezeptives sensorisches Feld mit exzitatorischer Wirkung auf das direkte Umfeld und inhibitorischer Wirkung auf weiter entfernte Bereiche. Während in einem linearen vollständig verknüpften Layer die Units völlig unabhängig voneinander sind, wird durch diese Implementierung einer reduzierten Aktivierung angrenzender Units der räumliche Zusammenhang zwischen benachbarten Units erzeugt.

### 10.3.1.3. Erweiterung der sprachlichen Stimuli

Neben dem visuellen Material haben wir auch das sprachliche Material erweitert. Zunächst einmal haben wir Präpositionen und Adverbien getrennt. Damit muss die Bedeutung der räumlich-relationalen Beschreibung kompositional aus verschiedenen Wörtern abgeleitet

<sup>27</sup>Entsprechende rezeptive Felder finden sich auch in der sensorischen Verarbeitung natürlicher neuronaler Systeme, etwa in der auditorischen oder der visuellen Verarbeitung (vgl. z.B. Eurich & Schwegler, 1997).

werden. Zudem werden die Satz­längen variabler und länger. Darüber hinaus haben wir auch einige neue Satz­typen eingeführt. Die Satz­typen ORO und ROO beschreiben wie gehabt die Positionierung eines Objekts in Relation zu einem anderen Objekt (Bsp. 34c und 34d). Die Satz­typen WOO und WRO fragen entweder das Objekt oder die Relation einer beschriebenen Szene ab (Bsp. 34e und 34f). Nach dem Fragezeichen (*QM*) setzen wir den Satz im Training mit der korrekten Antwort fort; in der Test-Phase können wir folglich überprüfen, ob das Netzwerk auch selbstständig die korrekte Antwort aus dem visuellen Input ableiten kann. Die Satz­typen OR und RO beschreiben die Position nur eines Objekts (Bsp. 34a und 34b). Allerdings werden wir auf diese Satz­typen nicht weiter eingehen.<sup>28</sup> Jede Situation lässt sich auch mit umgekehrter Relation beschreiben. Zudem kann das jeweils andere Objekt oder ein drittes Objekt als Ausgangspunkt der Beschreibung dienen. Somit kann eine Situation immer auf vielfach verschiedene Weise beschrieben werden.

- (34)
- a. BOS ObjektA ist rechts unten EOS (*Satztyp: OR*)
  - b. BOS rechts unten ist ObjektA EOS (*Satztyp: RO*)
  - c. BOS ObjektA ist rechts unter ObjektB EOS (*Satztyp: ORO*)
  - d. BOS rechts unter ObjektB ist ObjektA EOS (*Satztyp: ROO*)
  - e. BOS Wo ist ObjektA von ObjektB QM rechts unter EOS (*Satztyp: WOO*)
  - f. BOS Was ist rechts unter ObjektB QM ObjektA EOS (*Satztyp: WRO*)

### 10.3.2. Simulationen

#### 10.3.2.1. Trainings- und Test-Sets

Wir verwenden für SynSem II verschiedene Trainings- und Test-Sets, um besser kontrollieren zu können, welche Aspekte des Netzwerk-Verhaltens auf welche Aspekte der Stimuli zurückzuführen sind. Dabei unterscheidet sich die Anzahl der jeweils gleichzeitig gegebenen Input-Objekte und die Art des Rauschens (Positions- und Relationsrauschen). Des Weiteren unterscheiden wir originale und ideale Test-Sets: Bei originalen Test-Sets verwenden wir Target-Aktivierungen, in denen die originale Input-Positionierung der Objekte erhalten bleibt. Bei idealen Test-Sets nehmen die Objekte in der Target-Aktivierung idealisierte Positionen an. Sollte eine selbstorganisierte Prototypisierung stattfinden, so erwarten wir, dass bei Input-Situationen mit verschobenen Objekten bei den Output-Situationen idealisierte Positionierungen abgebildet werden. Eine dritte Unterscheidung betrifft spezielle Trainings- und Test-Sets, in denen im Training spezielle Relationen vorenthalten werden, für die im Test-Set die Performanz überprüft wird. Dies soll eine bessere Kontrolle der Systematizität der Generalisierung ermöglichen. Wir erläutern die Details der Manipulationen im jeweiligen Ergebnis-Abschnitt und unterscheiden zunächst die folgenden Trainings-Varianten:

---

<sup>28</sup>Da das Projekt SynSem II vor allem darauf abzielt, zu überprüfen, ob die Leistungen von SynSem I auch auf rein auto-assoziative Weise erbracht werden können, haben wir in der weiteren Darstellung auf die Untersuchung der Satz­typen OR und RO verzichtet. In den vorgestellten Simulationen waren diese Satz­typen nicht mehr im Trainings-Set enthalten. Die Fragesätze WOO und WRO haben wir in erweiterten Simulationen eingeführt, um die relationalen Satz­typen ORO und ROO syntaktisch zu variieren.

- Variante A mit Positionsrauschen: In Version A.a sind während des Trainings immer nur zwei Input-Objekte gegeben, in Version A.b sind immer drei Input-Objekte gegeben. Beide Versionen werden mit Situationen mit zwei und drei Objekten getestet.
- Variante B mit verschiedenen Anteilen an Feld-Abweichung: In Version B.a 10%, in Version B.b 30% und in Version B.c 50%. Die Versionen von Variante B wurden immer nur mit zwei Input-Objekten trainiert.
- Variante C: Kontrollierte Tests einer speziellen Relation, die im Trainings-Set vorenthalten wurde. In vier verschiedenen Versionen haben wir zunehmende Komplexität erzeugt, indem eine unterschiedliche Anzahl an Bedingungen für die Objekte SLASH und BACKSLASH vorenthalten wurde: in Version C.a eine Relation, in Version C.b zwei Relationen, in Version C.c drei Relationen und schließlich in Version C.d vier Relationen.

Die Trainings-Sets enthalten jeweils 100.000 Situationen mit korrespondierendem Satz, die Test-Sets enthalten jeweils 200 Satz-Situations-Paare. Wiederum sind Trainings- und Test-Sets so aufgebaut, dass sie zu jeweils 50% komplette Satz-Situations-Paare enthalten und zu jeweils 25% entweder nur visueller Input oder nur sprachlicher Input gegeben ist, um die verschiedenen Test-Modi Verstehen, Produktion und situiertes Verstehen zu untersuchen. Die Reihenfolge der Situations-Darbietung erfolgt mit vollständiger Permutation.

### 10.3.2.2. Parameter

Durch die auto-assoziative Realisierung und die damit verbundene Vergrößerung der Retina sowie des Outputs der visuellen Route mussten die internen Layer entsprechend angepasst werden. Wir verwenden in den Simulationen die folgenden Spezifikationen für die Architektur: Input- und Output-Layer der visuellen Route enthalten jeweils 289 Units, die Input- und Output-Layer der sprachlichen Route umfassen jeweils 12 bzw. 15 Units (abhängig davon, ob mit oder ohne Fragesätze trainiert wird). Der Layer für die rezeptiven Felder für die Vorverarbeitung der visuellen Stimuli umfasst 361 Units. Die beiden Kompressions-Layer der visuellen Route erhalten jeweils 68 Units, der rekurrente Integrations-Layer 136 Units. Für das Training verwenden wir wieder den Algorithmus Backpropagation Through Time und eine Batch-Größe von 1. In Übereinstimmung zum Vorgänger-Projekt haben wir für die Lernrate den Wert  $\epsilon = 0.2$  und für das Momentum den Faktor *momentum* = 0.6 gewählt. Nur die initiale Randomisierung der Gewichte wurde an die Größen-Zunahme angepasst, wir verwenden Werte innerhalb einer Obergrenze von *initialweightrange* = 0.2. Wir haben die Netzwerke jeweils über 20 Epochen trainiert. Grundlage der Analyse wurden wiederum die Ergebnisse für Epoche 15.

### 10.3.2.3. Präsentation der Ergebnisse

Anders als für die Ergebnis-Abschnitte für SynSem I werden wir die Ergebnisse im Folgenden nicht anhand der Test-Modi unterteilen. Stattdessen werden wir die Ergebnisse getrennt nach den drei Varianten präsentieren, die wir über die spezifische Gestaltung der Trainings- und Test-Sets unterschieden haben. Wir schreiten von Variante A (Positionsrauschen) zu

Variante B (Relationsrauschen) und schließen mit Variante C (Auslassungs-Tests). Dabei besprechen wir für jede Variante die Test-Modi der situierten und der entkoppelten Verarbeitung. Zudem werden wir die Unterschiede der Variationen innerhalb der Varianten thematisieren und auch variantenübergreifende Vergleiche anstellen. In einem vierten Unterkapitel besprechen wir für alle Varianten gemeinsam die Sprachproduktion. Zur Darstellung der Performanz verwenden wir die Kennwerte, die wir für die Ergebnisse der Simulation SynSem I vorgestellt haben (vgl. Kap. 10.2.2). Allerdings bestehen mit drei Varianten, drei Test-Modi und mitunter speziellen Test-Sets sowie kompletter und positionsabhängiger Analyse sehr viele mögliche Vergleichsmomente. Wir beschränken uns deshalb auf aussagekräftige und nicht-redundante Kennwerte und verzichten im Weiteren auf die Fehlerwerte EukDist Syn, EukDist Sem und PosRel Sem.

### 10.3.3. Ergebnisse Variante A: Positionsrauschen

Mit Variante A untersuchen wir zunächst einmal, wie unsere Architektur mit Rauschen im visuellen Input umgeht und ob das Netzwerk hierfür prototypisierte Objekt-Lokalisierungen herausbildet. Darüber hinaus untersuchen wir aber noch zwei weitere Aspekte: In Version A.a des Trainings-Sets verwenden wir ausschließlich Input-Situationen mit nur zwei Objekten. Die Performanz des Netzwerks wird anschließend mit Test-Sets überprüft, in denen die visuellen Input-Situationen über zwei oder drei Input-Objekte verfügen. Folgende Überlegung steckt hinter diesem Versuch: Aufmerksamkeit könnte in unserem Modell auch als Unterdrückung der Abbildung von Objekten verstanden werden, die in einer aktuellen Beschreibung irrelevant sind. Da das Netzwerk in Version A.a mit nur jeweils zwei Input-Objekten trainiert wird, erleichtern wir ihm diese Aufgabe: Es muss nur erkennen, welches Objekt im Vergleich zu den situationalen Konstellationen im Training irrelevantes Rauschen darstellt. Dies wird in der Variante mit Positionsrauschen zusätzlich dadurch erleichtert, dass die Objekte einer Relation immer in gegenüberliegenden retinalen Feldern erscheinen – ein drittes Objekt erscheint hingegen auf einem Feld, das diese Bedingung nicht erfüllt, und kann darüber als irrelevant identifiziert werden. In Version A.b verwenden wir ein Trainings-Set, in dem immer drei Input-Objekte gegeben sind. Diese Version dient der Kontrolle: Wir können damit überprüfen, ob auch dann eine Reduzierung der Objekt-Menge durch Unterdrückung stattfindet, wenn das Netzwerk auf die Abbildung von drei Objekten trainiert wird. Wie wir demonstrieren werden, erfüllen die Netzwerke unter den Bedingungen von Version A.a unsere Erwartungen, während die Unterdrückung der Objekt-Abbildung unter den Bedingungen von Version A.b nicht gelingt und hier auch die allgemeine Performanz schlechter ausfällt. Die allgemeine Performanz ist mit Einschränkungen für beide Versionen zufriedenstellend.

#### 10.3.3.1. Entkoppeltes Sprachverstehen (Variante A)

Wir beginnen mit dem entkoppelten Sprachverstehen, bei dem nur sprachlicher Input gegeben ist, und legen zunächst die Kennwerte für alle Trainings-Versionen dar.

**Globale Kennwerte: Visuelle Route** Für die positionsübergreifenden Fehlerwerte zeigt sich ein enttäuschendes Bild: Die Objekt-Lokalisierung mit  $\bar{x}_{\text{LocScore Sem}} = 0.364$  ( $\sigma =$

0.227), die Objekt-Identifizierung mit  $\bar{x}_{1-ObjCor Sem} = 0.514$  ( $\sigma = 0.339$ ) und der Gesamtfehler der visuellen Route mit  $\bar{x}_{1-Cos Sem} = 0.526$  ( $\sigma = 0.247$ ) deuten eine sehr schlechte Verarbeitungsleistung an. Auf der Satzposition *EOS* zeigt sich eine Verbesserung der Werte, die darauf hinweist, dass der sprachliche Input durchaus für die Konstruktion einer visuellen Situation genutzt wird:  $\bar{x}_{LocScore Sem} = 0.208$  ( $\sigma = 0.183$ ),  $\bar{x}_{1-ObjCor Sem} = 0.294$  ( $\sigma = 0.362$ ) und  $\bar{x}_{1-Cos Sem} = 0.361$  ( $\sigma = 0.266$ ). Allerdings sind diese Werte für die Position, auf der alle inhaltlichen Informationen gegeben sind, deutlich schlechter als entsprechende Werte in der Simulation SynSem I.

Klarheit über die Ursache dieser scheinbar schlechten Werte ergeben die folgenden beiden Überprüfungen: Erstens ein Vergleich von Version A.a (Training mit nur zwei Input-Objekten) mit Version A.b (Training immer mit drei Input-Objekten), und zweitens ein Vergleich von Ergebnissen für ein allgemeines Test-Set (d.h. mit verrauschten Objekt-Positionen) mit Ergebnissen für ein Test-Set mit idealen Objekt-Positionen. Für Version A.a, bei der im Training immer nur zwei Input-Objekte gegeben sind, ergeben sich folgende Werte: Positionsübergreifend zeigt sich nur eine leichte Verringerung des Fehlers mit  $\bar{x}_{LocScore Sem} = 0.3$  ( $\sigma = 0.252$ ),  $\bar{x}_{1-ObjCor Sem} = 0.374$  ( $\sigma = 0.33$ ) und  $\bar{x}_{1-Cos Sem} = 0.427$  ( $\sigma = 0.291$ ). Die Werte für die Satzposition *EOS* hingegen sind deutlich besser als die zuvor für alle Versionen zusammen betrachteten: Die Objekt-Lokalisierung erfolgt mit einem Fehler von  $\bar{x}_{LocScore Sem} = 0.106$  ( $\sigma = 0.132$ ) bereits sehr gut, die Objekt-Identifizierung ist mit  $\bar{x}_{1-ObjCor Sem} = 0.0$  perfekt. Es ergibt sich ein übergreifender Fehler von  $\bar{x}_{1-Cos Sem} = 0.178$  ( $\sigma = 0.236$ ). Sehr gut werden die Werte schließlich für die Test-Sets mit idealen Objekt-Positionen. Für diese ergibt sich kein Unterschied hinsichtlich des Inputs – es wird weiterhin nur sprachlicher Input gegeben. Allerdings werden die Objekte in diesen Test-Sets nicht auf irgendeiner Position, sondern auf den Ideal-Positionen (ohne Abweichungen) erwartet und für diese Positionierungen überprüft. Positionsübergreifend zeigt sich eine ähnliche Performanz mit  $\bar{x}_{LocScore Sem} = 0.237$  ( $\sigma = 0.277$ ),  $\bar{x}_{1-ObjCor Sem} = 0.38$  ( $\sigma = 0.332$ ) und  $\bar{x}_{1-Cos Sem} = 0.33$  ( $\sigma = 0.305$ ). Die Werte für die Satzposition *EOS* mit vollständiger Information schließlich weisen eine Performanz auf, die mit den Ergebnissen aus Simulation SynSem I (vgl. Kap. 10.2.3.1) vergleichbar ist:  $\bar{x}_{LocScore Sem} = 0.003$  ( $\sigma = 0.003$ ),  $\bar{x}_{1-ObjCor Sem} = 0.0$  und  $\bar{x}_{1-Cos Sem} = 0.009$  ( $\sigma = 0.01$ ). Eine ANOVA für den Fehlerwert 1-Cos Sem auf der Satzposition *EOS* zeigt für das Test-Set mit idealen Positionierungen einen signifikanten Unterschied zwischen den beiden Versionen A.a und A.b ( $df = 2, F = 1022.1, p < 0.001$ ).<sup>29</sup> Aus den Ergebnissen können wir schließen, dass die Version mit reduziertem Input wesentlich erfolgreicher war und dass beim Sprachverstehen Situationen mit prototypisierten Objekt-Positionierungen konstruiert werden.

**Globale Kennwerte: Sprachliche Route** Für die Wortvorhersage-Komponente zeigen sich Verhältnisse, die mit den Ergebnissen für SynSem I übereinstimmen. Wir verwenden wieder die gemittelten Kennwerte für alle Trainings-Versionen: Während die kategoriale Vorhersage in positionsübergreifender Betrachtung mit dem Wert  $\bar{x}_{1-CatCor Syn} = 0.077$  ( $\sigma = 0.267$ ) sehr gut ausfällt, wirkt sich der fehlende visuelle Input auf den übergreifenden Fehler aus. Für diesen wird nur ein Wert von  $\bar{x}_{1-Cos Syn} = 0.304$  ( $\sigma = 0.335$ ) erreicht, da

<sup>29</sup>Wir verwenden für die ANOVA die Funktion *anova* auf Grundlage eines linearen Modells, das mit der Funktion *lm* berechnet wurde. Beide Funktionen sind im Standard-R-Package *stats* verfügbar (R Development Core Team, 2010).

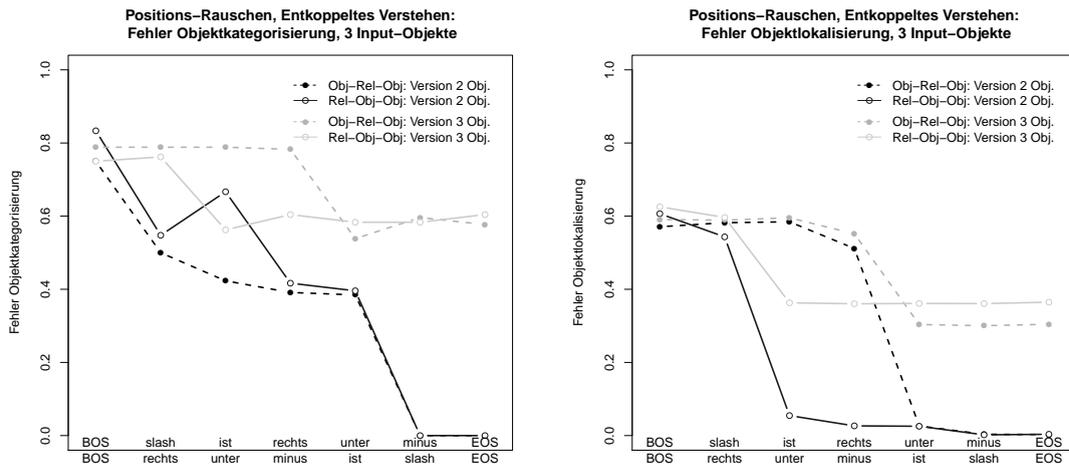
sich keine situational-semantischen Einschränkungen ergeben. Für die Ergebnisse der unterschiedlichen Trainings- und Test-Versionen ergeben sich weitgehend identische Fehlerwerte. Damit zeigen sich für die Performanz der sprachlichen Route im Unterschied zur visuellen Route keine Auswirkungen der zugrundegelegten Version. Auch das verwendete Test-Set wirkt sich nicht auf die Performanz aus: Die Ergebnisse sind deckungsgleich, da der sprachliche Input für ideale und abweichende Objekt-Positionierungen identisch ist. Schließlich ziehen wir wieder den Fehler für die einzelne Satzposition des *EOS* heran: Für diese zeigt sich wie erwartet ein minimaler Fehler von  $\bar{x}_{1-CatCor Syn} = 0.0$  und  $\bar{x}_{1-Cos Syn} < 0.001$  ( $\sigma < 0.001$ ). Dieser Minimalfehler zeigt sich auch in allen anderen Bedingungen und Varianten von SynSem II beim entkoppelten und situierten Verstehen, weshalb wir im Weiteren auf seine Darstellung verzichten werden.

**Detailanalyse: Visuelle Route** In Abbildung 10.19 werden die Fehler der Identifizierung (Abb. 10.19(a)) und der Lokalisierung (Abb. 10.19(b)) der Objekte sowie der Gesamtfehler der Performanz der visuellen Route (Abb. 10.19(c)) dargestellt. Zunächst einmal zeigt sich wie aus den allgemeinen Kennwerten erwartet der Unterschied zwischen Version A.a und A.b, der sich in einer wesentlich geringeren Abnahme des Fehlers für Version A.b äußert. Zugleich zeigt sich für diese Version, dass auch hier eine Entwicklung stattfindet, die von den Informationen des visuellen Inputs geleitet wird. Während der Fehler hier aber für alle drei Kennwerte auf einem hohen Niveau verbleibt, erreicht er in Version A.a mit dem jeweiligen Erlangen der vollständigen Information für Lokalisierung und Identität einen minimalen Wert. Die Verbesserung bei der Lokalisierung (LocScore Sem) ist verständlicherweise in beiden Versionen direkt von den Adverbien abhängig, wobei die wesentlich größere Fehlerminimierung mit dem zweiten Adverb einhergeht, da erst die kombinierte Information die relativen Positionen der Objekte exakt bestimmt.<sup>30</sup> Aufgrund der Trennung der Adverbien sowie der Präpositionen in der Grammatik von SynSem II können wir damit wiederum zeigen, dass das Netzwerk die Kombination sequenzieller Informationen unmittelbar und auch für kleinere Sequenzabschnitte für die Situationskonstruktion nutzen kann. Interessant ist der Unterschied zwischen den Versionen beim Entwicklungsverlauf für die Objekt-Identifizierung (1-ObjCor Sem): Für Version A.a zeigen sich die Verbesserungen jeweils beim Erscheinen der Nomen, die die relevanten Objekte benennen, während sich für die anderen Wörter nur geringfügige Verbesserungen zeigen – oder im Fall des Satztyps ROO auf Position 3 (Adverb/Präposition) sogar eine Verschlechterung. Für Version A.b hingegen können wir beobachten, dass die Verbesserung für die Objekt-Identifizierung der Situationsbeschreibung praktisch überhaupt nicht von den Informationen der Nomen abhängt, sondern direkt durch die Adverbien bestimmt wird. Dabei können diese aber nur eine Einschränkung der Aktivierungsverteilung auf die relevanten Positionen erbringen, eine Objekt-Identifizierung scheint in dieser Version nur vage zu erfolgen.

**Detailanalyse: Sprachliche Route** Abbildung 10.20 zeigt den Fehlerverlauf der Wortvorhersage für die beiden Versionen A.a und A.b anhand des Gesamtfehlers 1-Cos Syn.

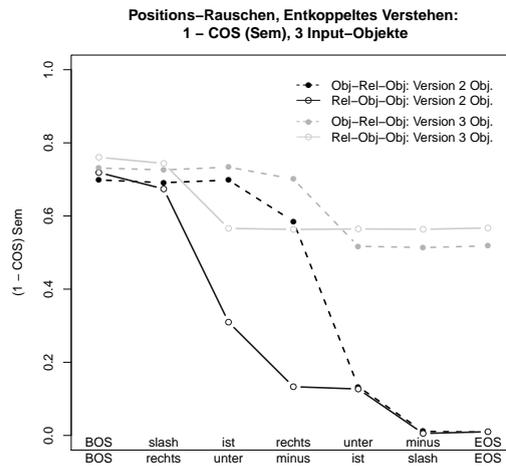
---

<sup>30</sup>Die leichte Verbesserung für das erste Adverb ist möglicherweise vor allem auf die Beschreibungen der Relationen ÜBER und UNTER zurückzuführen, in denen auf das Adverb keine Präposition bzw. kein zweites Adverb folgt. Hier ist alle Information über die Lokalisierung bereits nach dem ersten Adverb gegeben.



(a) Objekt-Identifizierung (1-ObjCor Sem)

(b) Objekt-Lokalisierung (LocScore Sem)



(c) Fehler Objekt-Relation (1-Cos Sem)

Abbildung 10.19.: SynSem II, Variante A: Performanz der visuellen Route beim entkoppelten Verstehen: Objekt-Identifizierung (1-ObjCor), Objekt-Lokalisierung (LocScore) und Gesamtfehler der Abbildung korrekter Objekt-Relationen (1-Cos Sem). Vergleich der Trainings-Versionen A.a (zwei Input-Objekte) und A.b (drei Input-Objekte).

Wie in SynSem I zeigen sich wieder die Auswirkungen des verwendeten Fehlers, der durch die Verwendung der Output- und Target-Vektoren (Aktivierung aller kategorial möglichen Wörter gegenüber der Erwartung nur einer aktivierten Target-Unit) zustandekommt: Wortpositionen mit einer Vorhersage mehrerer möglicher Folgewörter erhalten einen höheren Fehler, da durch die Verteilung der Aktivierung über alle möglichen Wörter die Aktivierung des Target-Worts geringer ausfällt. Eine Fehlerminimierung zeigt sich hingegen auf den Positionen des Verbs sowie für emphEOS und emphBOS. Da die sprachliche Verarbeitung nicht durch den situationalen Kontext gestützt ist, kann keine Einschränkung möglicher Wörter erfolgen. Einzige Ausnahme ist eine leichte Fehlerverringering für das jeweils zweitgenannte Objekt: Hier kann sich der sequenziell-sprachliche Kontext auswirken, denn da eine Situation immer zwei unterschiedliche Objekte enthält, kann auf dieser Position die Aktivierung des erstgenannten Objekts ausgeschlossen werden. Der Fehler verläuft für die beiden Trainings-Versionen innerhalb der Satztypen weitgehend deckungsgleich. Die einzigen Abweichungen zeigen sich für den Satztyp ROO auf den Positionen 1 und 3. Der erhöhte Fehler für Version A.a auf Position 3 beruht möglicherweise darauf, dass in allen Sätzen im Test-Modus des entkoppelten Verstehens das *minus* das am stärksten aktivierte Nomen ist, was möglicherweise auf eine leichte distributionale Varianz im Trainingskorpus zurückgeführt werden kann.<sup>31</sup> Auf die Abbildung des kategorialen Vorhersagefehlers verzichten wir an dieser Stelle, da ab dem Verb der Fehler ab Satzposition 2 in allen Bedingungen (Trainings-Version, Test-Version, Satztyp) auf 0.0 fällt. Problematisch für die kategoriale Vorhersage ist also wie bereits in Simulation SynSem I nur die satzinitiale Position, auf der bei noch unbekanntem Satztyp gleichermaßen Nomen und Adverbien erscheinen können.

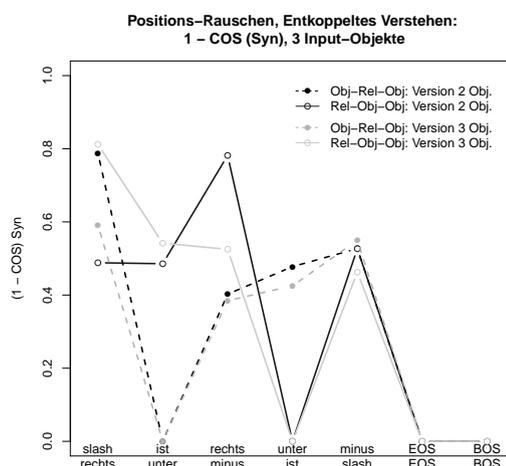


Abbildung 10.20.: SynSem II, Variante A: Ergebnisse für die Wortvorhersage-Performanz beim entkoppelten Verstehen (1-Cos Syn). Vergleich der Trainings-Versionen A.a (zwei Input-Objekte) und A.b (drei Input-Objekte).

<sup>31</sup>Die Aktivierungsunterschiede könnten auf leichte Ungleichmäßigkeiten der Auftretenswahrscheinlichkeit der Nomen im Trainingskorpus zurückzuführen sein. In der Tat tritt *minus* mit einer Häufigkeit von 25.11% etwas öfter auf als die anderen drei Nomen mit 25.04%, 24.97% und 24.88%. Allerdings ist diese Erklärung bei solch marginalen Unterschieden nicht sehr wahrscheinlich.

**Zusammenfassung** Insgesamt demonstrieren die Ergebnisse für Variante A beim entkoppelten Sprachverstehen sowohl für die visuelle als auch die sprachliche Verarbeitung Performanzmerkmale mit weitgehender Übereinstimmung zu den Merkmalen, die wir für den entsprechenden Testmodus für die Simulation SynSem I beobachten konnten (vgl. Kap. 10.2.3.1). Eine wichtige Information liefert uns die Feststellung, dass die Objekt-Identifizierung in Version A.a durch die Indizierung der entsprechenden Situations-Konstruktion über die Nomen bestimmt wird, in Version A.b hingegen nur die Adverbien eine Verbesserung erbringen. Und auch die Nutzung der Informationen für die Lokalisierung durch die Adverbien erfolgt nur beschränkt. Wie wir gleich für das situierte Verstehen zeigen werden, ist die Objekt-Identifizierung und -Lokalisierung auf Grundlage des Verarbeitens des visuellen Inputs hingegen nicht beeinträchtigt. Das Training mit jeweils drei Input-Objekten scheint eine Erschwernis mit sich zu bringen, Objekte an Nomen und Lokalisierungen an Adverbien und Präpositionen zu binden und damit Situationsaspekte sprachlich zu indizieren. Aus der demgegenüber erfolgreich verlaufenden Indizierung bzw. dem erfolgreichen Re-Enactment für Version A.a können wir schließen, dass die Reduzierung der Objektmenge – vergleichbar mit einer geleiteten oder geteilten Aufmerksamkeit und einer Fokussierung auf einen beschränkten Ausschnitt einer Situation – unmittelbar relevant ist für den Bedeutungserwerb.

### 10.3.3.2. Situierte Sprachverarbeitung (Variante A)

Im Unterschied zur entkoppelten Verarbeitung spielt es für die situierte Verarbeitung weder eine Rolle, ob Trainings-Version A.a oder A.b verwendet wird, noch ob die Test-Sets mit idealen oder verrauschten Positionen verwendet werden. Die Fehlerwerte sind unter allen Bedingungen weitgehend deckungsgleich. Sowohl für die visuelle als auch die sprachliche Route zeigen sich bereits im positionsübergreifenden Vergleich sehr gute Werte.

**Globale Kennwerte: Visuelle Route** Die Fehler für die Objekt-Lokalisierung und für die Objekt-Identifizierung liegen mit  $\bar{x}_{\text{LocScore Sem}} = 0.089$  ( $\sigma = 0.134$ ) und  $\bar{x}_{1\text{-ObjCor Sem}} = 0.158$  ( $\sigma = 0.243$ ) in der gleichen Größenordnung wie in SynSem I. Der Fehlerwert für den gesamten visuellen Output ist sogar geringfügig besser mit  $\bar{x}_{1\text{-Cos Sem}} = 0.142$  ( $\sigma = 0.203$ ). Betrachten wir allerdings die Werte für die Satzposition *EOS*, auf der die vollständigen sprachlichen Informationen gegeben sind, zeigt sich, dass keinerlei Veränderung eingetreten ist:  $\bar{x}_{\text{LocScore Sem}} = 0.092$  ( $\sigma = 0.14$ ),  $\bar{x}_{1\text{-ObjCor Sem}} = 0.159$  ( $\sigma = 0.246$ ) und  $\bar{x}_{1\text{-Cos Sem}} = 0.144$  ( $\sigma = 0.205$ ). Es erfolgt also keinerlei Verbesserung durch den sprachlichen Input. Dies deutet auf zwei Unterschiede zu SynSem I hin (vgl. hierzu Kap. 10.2.3.2): Erstens scheint der sprachliche Input aus bestimmten Gründen keinen positiven Einfluss auf die visuelle Verarbeitung nehmen zu können. Zweitens erreicht die visuelle Verarbeitung bei vollständiger Information ein geringeres Niveau. Die Tatsache, dass die Test-Sets mit Verwendung idealer Objekt-Positionierung keinen Unterschied erbringen, weist auf eine weitere Eigenschaft hin: Es scheint keine Prototypisierung der Objekt-Lokalisierung zu erfolgen, die visuelle Verarbeitung bildet schlichtweg die Verhältnisse des Inputs ab, unabhängig davon, ob die Objekte ideal oder versetzt positioniert sind.

**Globale Kennwerte: Sprachliche Route** Für die sprachliche Route deuten sich die gleichen Gegebenheiten an, die wir schon für SynSem I beobachten konnten: Der Fehler der kategorialen Vorhersage liegt bei  $\bar{x}_{1-CatCor Syn} = 0.077$  ( $\sigma = 0.266$ ) und damit gleichauf mit der Performanz beim entkoppelten Sprachverstehen. Der Gesamtfehler hingegen, in den wiederum auch semantische Gesichtspunkte eingehen, ist mit  $\bar{x}_{1-Cos Syn} = 0.137$  ( $\sigma = 0.291$ ) wesentlich besser. Wir können also festhalten, dass die Informationen des visuellen Inputs für die lexikalische Antizipation bei der Wortvorhersage genutzt werden.

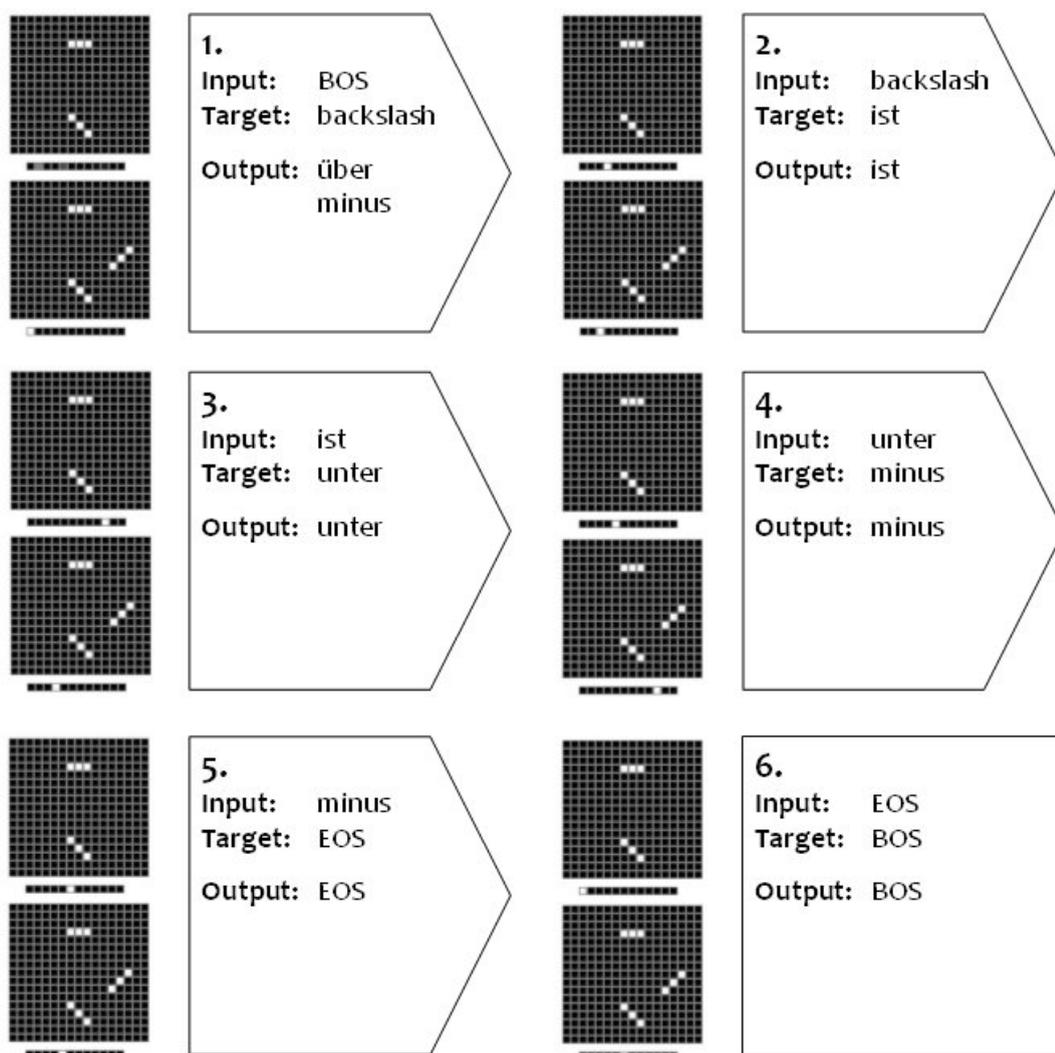


Abbildung 10.21.: SynSem II, Variante A.a (Training mit 2 Objekten): Beispiel zu situiertem Verstehen bei drei Input-Objekten.

**Beispiele: Variante A.a und Variante A.b** Die folgenden beiden Beispiele (Abb. 10.21 und 10.22) demonstrieren die situierte Sprachverarbeitung für SynSem II in Variante A und verdeutlichen den Unterschied zwischen den beiden Versionen A.a und A.b bei der Verarbeitung von Sätzen bei einem situationalen Kontext mit drei Input-Objekten. Das Beispiel in Abbildung 10.21 zeigt die Aktivierungs-Details für die Version A.a bei der Verarbeitung des Satzes *BOS backslash ist unter minus EOS*. Im visuellen Input sind

die Objekte BACKSLASH, SLASH und MINUS gegeben. Da Objekte in Variante A nur dann in einer aufeinander bezogenen räumlichen Konstellation stehen können, wenn sie gegenüberliegende Felder einnehmen, ist MINUS ÜBER BACKSLASH die einzige valide Objekt-Relation der gegebenen Situation.

1. Im Output der visuellen Route zeigt sich von Beginn an die korrekte Objekt-Relation. Folglich werden auch nur zu dieser Relation korrespondierende Wörter vorhergesagt, *minus* und *über*.
2. Mit dem Input des Wortes *backslash* können zusammen mit dem visuellen Kontext auch alle sprachlichen Informationen für Satztyp und Folgewörter antizipiert werden. Die Informationslage entspricht der Bedingung in SynSem I mit nur zwei Input-Objekten. Dementsprechend wird korrekt ausschließlich das Verb *ist* vorhergesagt.
3. - 6.: Es zeigen sich nur noch korrekte Vorhersagen jeweils eines Wortes. Auch der visuelle Output bleibt konstant korrekt.

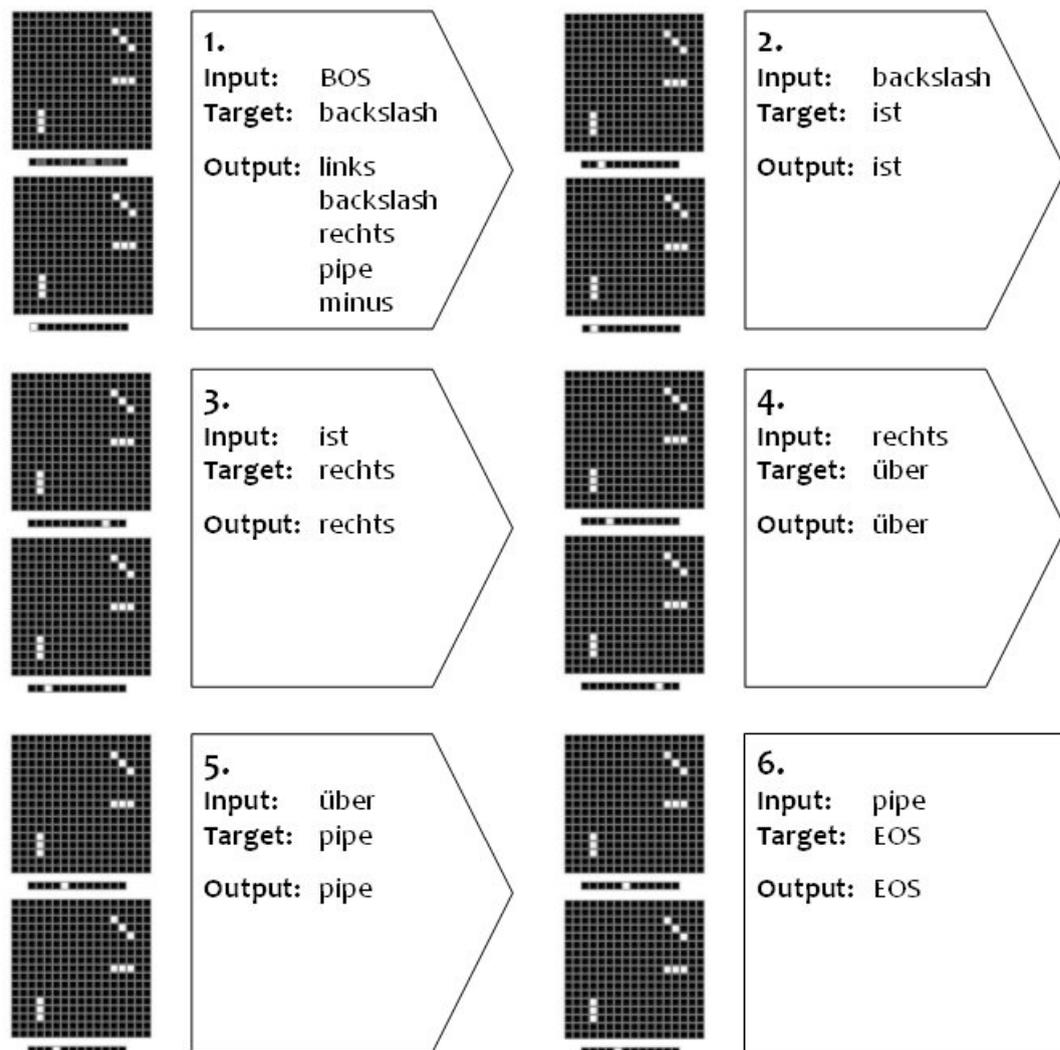


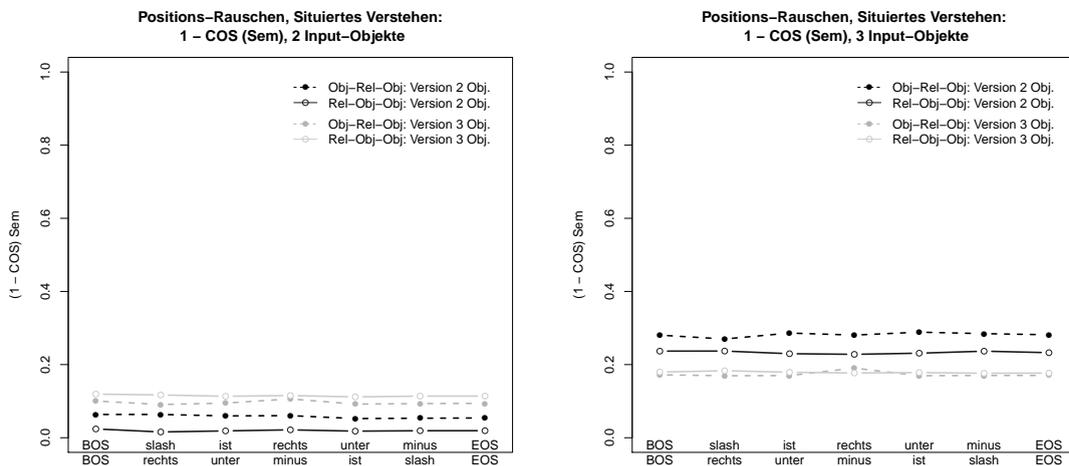
Abbildung 10.22.: SynSem II, Variante A.b (Training mit drei Objekten): Beispiel zu situiertem Verstehen bei drei Input-Objekten.

Das Beispiel in Abbildung 10.22 stellt dem die Aktivierungs-Details für die Version A.b bei der Verarbeitung des Satzes *BOS backslash ist rechts über pipe EOS* gegenüber. Im visuellen Input sind wiederum drei Objekte gegeben: BACKSLASH, PIPE und MINUS. Wiederum erfüllt nur eine Objekt-Relation das Positionierungs-Kriterium: BACKSLASH RECHTSÜBER PIPE. Das Objekt MINUS hingegen ist für die Situation irrelevant. Dennoch wird es (den Bedingungen des Trainings in Version A.b entsprechend) dargestellt – im visuellen Output finden sich folglich konstant alle drei Objekte in ihrer Ausgangspositionierung. Auch nach vollständig gegebener sprachlicher Information ergibt sich keine Reduzierung des visuellen Outputs.

1. Neben der korrekten Abbildung aller drei Input-Objekte im visuellen Output werden im Output der sprachlichen Route alle Wörter aktiviert, die in der aktuellen Situation zugegen sind. Auch die Benennung des hinsichtlich seiner Positionierung eigentlich ausgeschlossenen Objekts MINUS wird aktiviert – allerdings vergleichsweise schwach.
2. Im Kontext des gegebenen visuellen Inputs sind mit dem sprachlichen Input des Wortes *backslash* wiederum alle erforderlichen Informationen gegeben, um alle Folgewörter antizipieren zu können. Dementsprechend wird korrekt ausschließlich das Verb vorhergesagt.
3. - 7.: Auf allen folgenden Positionen wird jeweils das einzige korrekte Wort vorhergesagt (Vorhersage von *BOS* auf Position 7 wird nicht gezeigt).

**Detailanalyse: Visuelle Route** Die Diagramme in Abbildung 10.23 zeigen die Auswirkungen der Trainings-Versionen A.a und A.b auf den Gesamtfehler 1-Cos Sem, wenn im Test-Set zwei Input-Objekte (Abb. 10.23(a)) oder drei Input-Objekte (Abb. 10.23(b)) gegeben sind. Auffällig ist, dass sich im Unterschied zu Simulation SynSem I praktisch keine Entwicklung des Fehlers ergibt. Dies demonstriert die oben erwähnte Unabhängigkeit der visuellen Verarbeitung vom sprachlichen Input, die darauf beruht, dass dieser keinerlei Zusatzinformation zur korrekten Verarbeitung im Sinne etwa von Aufmerksamkeitslenkung erbringt. Die Auswahl der relevanten Konstellation wird alleine durch das Positionierungs-Kriterium erfüllt. Bei zwei Input-Objekten scheint sich für Version A.a eine marginal bessere Performanz zu ergeben. Bei drei Input-Objekten schneiden beide Versionen etwas schlechter ab. Zudem zeigt sich hier ein leichter Vorteil für Version A.b, der darauf zurückzuführen ist, dass in Version A.a eine Reduzierung stattfindet, während mit dem Target-Vektor (auf dessen Vergleich der Fehlerwert beruht) drei Output-Objekte erwartet werden.

**Detailanalyse: Sprachliche Route** Die Diagramme in Abbildung 10.24 zeigen den Gesamtfehler der Wortvorhersage (1-Cos Syn). Innerhalb der Diagramme vergleichen wir wieder die Performanz für die Versionen A.a und A.b in Abhängigkeit des Satztyps. Über die Diagramme hinweg vergleichen wir die Performanz bei zwei Input-Objekten (Abb. 10.24(a)) gegenüber drei Input-Objekten im Test-Set (Abb. 10.24(b)). In den Abbildungen für die unterschiedliche Anzahl der Input-Objekte zeigt sich wie erwartet ein deutlicher Unterschied zur Performanz im Testmodus mit entkoppeltem Verstehen (vgl. Abb. 10.20): Beim situierten Verstehen wird der situationale Kontext für die semantische Einschränkung

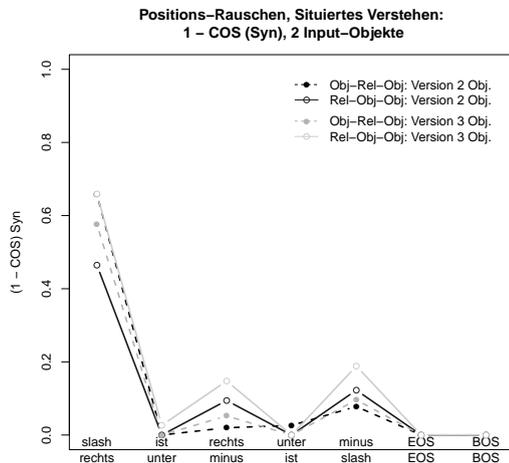


(a) Fehler Objekt-Relation (1-COS Sem) 2 Objekte (b) Fehler Objekt-Relation (1-COS Sem) 3 Objekte

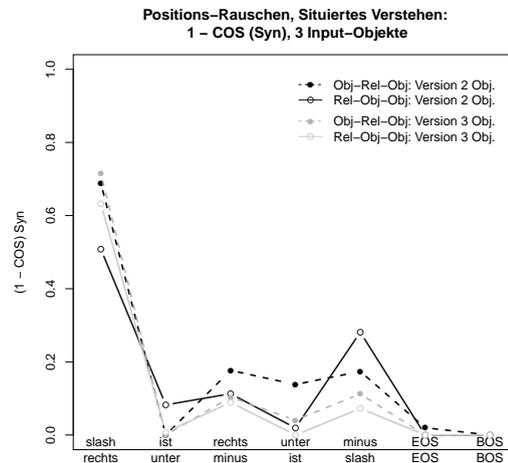
Abbildung 10.23.: SynSem II, Variante A: Ergebnisse für die visuelle Verarbeitung bei situierter Sprachverarbeitung anhand des Vergleichs der Ähnlichkeit des Aktivierungs- und Target-Vektors (1-COS Sem). Innerhalb der Diagramme: Vergleich der Performanz der Trainings-Versionen A.a und A.b. Zwischen den Diagrammen: Vergleich der Performanz bei zwei und drei Input-Objekten im Test-Set.

möglicher Wörter und eine dementsprechend präzisere Vorhersage der Folgewörter genutzt. Allerdings erfolgt diese Präzisierung nicht perfekt, was zum Beispiel auf Position 5 ersichtlich wird, denn es wäre zu erwarten gewesen, dass der Fehler hier auf einen minimalen Wert zustrebt, da alle Informationen für eine ausschließliche Vorhersage des Namens des zweiten Objekts verfügbar sind. Während Version A.a bei nur zwei Input-Objekten einen leichten Vorteil zu haben scheint, dreht sich diese Lage bei drei Input-Objekten zugunsten von Version A.b (vgl. v.a. Position 3 und 5). Insbesondere der vergleichsweise hohe Fehler von Version A.a bei drei Input-Objekten auf Position 5 für das zweitgenannte Objekt weist darauf hin, dass das dritte Objekt im visuellen Input zu Interferenzen bei der Objekt-Benennung und einer etwas abgeschwächten Aktivierung des korrekten Nomens führt. Insgesamt betrachtet zeigt sich aber eine sehr gute Performanz mit erfolgreicher Nutzung des situationalen Kontexts, die sich mit der Qualität der korrespondierenden Performanz von SynSem I vergleichen lässt (vgl. Abb.10.12(a)).

In Abbildung 10.25 vergleichen wir die Performanz für die Fragesätze in Version A.a bei zwei Input-Objekten im Test-Set. Wiederum ergeben sich für den Kennwert 1-Cos Syn teilweise erhöhte Fehlerwerte auf den Positionen 3 und 5 (vgl. Abb. 10.25(a)). Für den Fragesatztyp mit *wo*-Frage bleibt der Fehler ab Position 4 minimal. Der niedrige Fehlerwert auf den Positionen nach dem Fragezeichen-Marker *QM* zeigen, dass die Fragen korrekt beantwortet werden, wobei die *was*-Frage geringfügig schwerer zu beantworten zu sein scheint. Abbildung 10.25(b) demonstriert, dass die kategoriale Vorhersage (1-CatCor) nach Position 2 unbeeinträchtigt verläuft. Die Erweiterung um die neuen Satztypen wirkt sich also nicht negativ auf die Performanz aus.

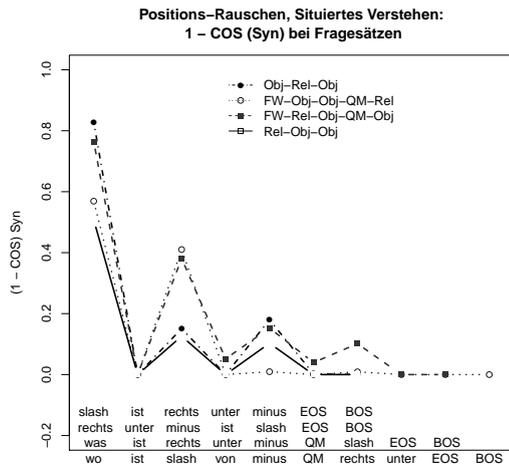


(a) Wortvorhersage (1-Cos Syn), 2 Obj.

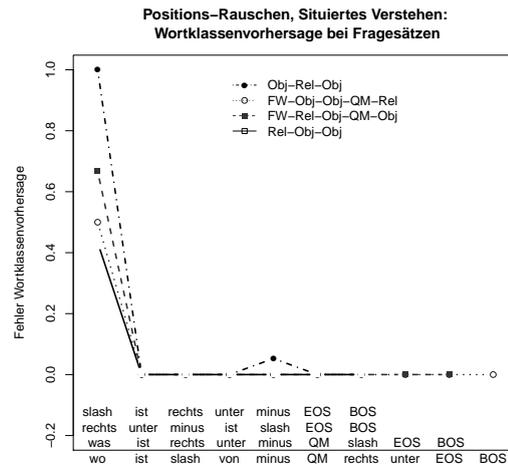


(b) Wortvorhersage (1-Cos Syn), 3 Obj.

Abbildung 10.24.: SynSem II, Variante A: Performanz der sprachlichen Route bei situierter Verarbeitung: Vergleich der Performanz für verschiedene Test-Situationen mit zwei Input-Objekten und drei Input-Objekten.



(a) Wortvorhersage (1-Cos Syn), Frage



(b) Wortvorhersage (1-CatCor), Frage

Abbildung 10.25.: SynSem II, Variante A: Performanz der sprachlichen Route bei situierter Verarbeitung für Fragesätze: Vergleich der Performanz für wo- und was-Fragen.

**Zusammenfassung** Die guten Ergebnisse für die Performanz in Version A zeigen, dass das Modell mit Rauschen (i.S.v. Verschiebungen der Objekte von Idealpositionen) umgehen kann. Beispiel 10.21 demonstriert, dass bei einer entsprechenden Gestaltung der Trainings-Stimuli eine automatische Reduzierung der Objektmenge erfolgt. Aufmerksamkeit wird hier als Unterdrückung der Abbildung von irrelevanten Objekten bzw. als Filterung irrelevanter Bestandteile komplexer Situationen simuliert. Allerdings erfolgt die Reduzierung nicht – wie von uns in Entsprechung zu Simulation SynSem I erwartet – durch den sprachlichen Input, sondern durch Informationen des visuellen Inputs. Dementsprechend lassen sich nur für den sprachlichen Output Inferenzen in Form von modalitätsübergreifendem Priming zeigen, nicht aber für den visuellen Output. Mit den Ergebnissen für das entkoppelte Sprachverstehen ergibt sich folgendes Bild: Ohne visuellen Kontext wird das Rauschen in einer Prototypisierung der Objekt-Positionierung umgesetzt, die Wortvorhersage erfolgt nur auf Grundlage der sprachlichen Informationen. Die erfolgreiche Indizierung der Situations-Konstruktion ist dabei direkt abhängig von der Form des Trainings – sie gelingt nur, wenn hier eine Reduzierung der relevanten Objekte vorgegeben wird. Mit visuellem Kontext erfolgt die visuelle Verarbeitung auf Grundlage des visuellen Inputs, der sprachliche Input hingegen nimmt in der Simulation SynSem II keinen Einfluss auf die Performanz der visuellen Route. Dementsprechend erfolgt hier auch keine Prototypisierung der Objekt-Positionierung – in entsprechenden Beispielen (hier nicht abgebildet) zeigt sich dies darin, dass die Objekte auf der verschobenen Position abgebildet werden, auf der sie im Input erscheinen. Umgekehrt wirkt sich der visuelle Kontext aber auf die Performanz der sprachlichen Route aus, für die wir Inferenzen im Sinne eines modalitätenübergreifenden Priming-Effekts beobachten können.

### 10.3.4. Ergebnisse Variante B: Relationsrauschen

In Variante B verwenden wir im Training nur Input-Situationen mit zwei Objekten, die aber mit variablem Anteil auf verschiedenen nicht idealen (d.h. nicht gegenüberliegenden) Positionen erscheinen können. Die abweichenden Positionierungen der Objekt-Relationen erscheinen dabei zusätzlich zum Positionsrauschen, das wie in Version A in 15% der Fälle eine Abweichung um eine Unit und in 5% der Fälle eine Abweichung um zwei Units beträgt. Neben der allgemeinen Performanz untersuchen wir, ob sich unter den gegebenen Bedingungen eine Prototypisierung der Objekt-Positionierung entwickelt und welchen Einfluss der Anteil abweichender relationaler Positionierungen (10%, 30% und 50%) auf die Performanz nimmt.

#### 10.3.4.1. Entkoppeltes Sprachverstehen (Variante B)

**Globale Kennwerte** Wir beginnen wieder mit dem entkoppelten Verstehen, bei dem sich das gleiche vermeintliche Problem zeigt, das wir bei der Variante mit Positionsrauschen beobachten konnten: Wenn wir als Grundlage die Test-Versionen mit versetzten Positionen – gemittelt über die Versionen mit 10%, 30% und 50% Abweichung – verwenden, zeigen sich vergleichsweise hohe positionsübergreifende Fehlerwerte mit  $\bar{x}_{\text{LocScore Sem}} = 0.405$  ( $\sigma = 0.285$ ),  $\bar{x}_{1\text{-ObjCor Sem}} = 0.35$  ( $\sigma = 0.282$ ) und  $\bar{x}_{1\text{-Cos Sem}} = 0.541$  ( $\sigma = 0.287$ ). Ziehen wir die positionsabhängigen Fehlerwerte auf der Satzposition *EOS* heran, auf der die vollständigen

sprachlichen Informationen gegeben sind, zeigt sich eine deutliche Verbesserung aller drei Fehlerwerte, die nun  $\bar{x}_{\text{LocScore Sem}} = 0.169$  ( $\sigma = 0.31$ ),  $\bar{x}_{\text{1-ObjCor Sem}} = 0.043$  ( $\sigma = 0.141$ ) und  $\bar{x}_{\text{1-Cos Sem}} = 0.36$  ( $\sigma = 0.325$ ) betragen. Interessanterweise zeigt sich die deutlichste Verbesserung für die Objektidentifikation, die ja unabhängig von ihrer Lokalisierung erfolgt. Die Performanz der sprachlichen Route zeigt beim entkoppelten Verstehen die gewohnten Verhältnisse: einen positionsübergreifend geringen kategorialen Vorhersagefehler von  $\bar{x}_{\text{1-CatCor Syn}} = 0.078$  ( $\sigma = 0.268$ ) und einen vergleichsweise hohen Gesamtfehler von  $\bar{x}_{\text{1-Cos Syn}} = 0.294$  ( $\sigma = 0.308$ ).

Ziehen wir für die Performanz der visuellen Route zum Vergleich die Ergebnisse für das Test-Set mit idealen Positionierungen heran, die zwar den gleichen sprachlichen Input bereitstellen, jedoch andere Target-Vektoren für den visuellen Output vorgeben: Hier sind die positionsübergreifenden Fehlerwerte bereits wesentlich geringer mit  $\bar{x}_{\text{LocScore Sem}} = 0.246$  ( $\sigma = 0.258$ ),  $\bar{x}_{\text{1-ObjCor Sem}} = 0.351$  ( $\sigma = 0.281$ ) und  $\bar{x}_{\text{1-Cos Sem}} = 0.359$  ( $\sigma = 0.284$ ). Auf der Satzposition *EOS* schließlich zeigen sich Fehlerwerte, die auf eine Performanz hinweisen, die mit SynSem I vergleichbar ist:  $\bar{x}_{\text{LocScore Sem}} = 0.029$  ( $\sigma = 0.031$ ),  $\bar{x}_{\text{1-ObjCor Sem}} = 0.06$  ( $\sigma = 0.163$ ) und  $\bar{x}_{\text{1-Cos Sem}} = 0.068$  ( $\sigma = 0.085$ ). Dies weist auf zwei Gegebenheiten hin: Erstens werden anhand des sprachlichen Inputs valide situationale Abbildungen konstruiert, und zweitens zeigt sich für diese Abbildungen eine Prototypisierung der Objekt-Positionierung. Eine weitere Verbesserung ergibt sich mit den verschiedenen Trainings-Versionen B.a, B.b und B.c, in denen verschiedene Anteile abweichender relationaler Positionierungen gegeben sind (10%, 30% und 50%). Wir ziehen zum Vergleich den Fehlerwert 1-Cos Sem auf der Satzposition *EOS* heran: In Version B.a zeigt sich der niedrigste Fehler mit  $\bar{x}_{\text{1-Cos Sem}} = 0.015$  ( $\sigma = 0.015$ ), in Version B.b ein deutlich höherer mit  $\bar{x}_{\text{1-Cos Sem}} = 0.079$  ( $\sigma = 0.059$ ) und der höchste Fehler in Version B.c mit  $\bar{x}_{\text{1-Cos Sem}} = 0.11$  ( $\sigma = 0.115$ ). Eine ANOVA für den Fehlerwert 1-Cos Sem auf dieser Satzposition weist auf einen signifikanten Unterschied zwischen den drei Versionen hin ( $df = 2, F = 21.059, p < 0.001$ ). Ein paarweise vergleichender t-Test mit Korrektur nach Holm zeigt, dass der Unterschied zwischen allen drei Versionen signifikant wird (B.a zu B.b:  $p < 0.001$ , B.a zu B.c:  $p < 0.001$ , B.b zu B.c:  $p < 0.05$ ).<sup>32</sup> Während sich diese Unterschiede auch in den anderen Fehlerwerten der visuellen Route finden, ergeben sich für die Fehlerwerte der sprachlichen Route keine signifikanten Unterschiede.

**Beispiel: Entkoppeltes Verstehen** Abbildung 10.26 zeigt die Aktivierungs-Details des Modells in der Version B.c im Test-Modus des entkoppelten Verstehens für den Satz *BOS rechts unter backslash ist slash EOS*. Im Beispiel zeigt sich zum einen die inkrementelle Verarbeitung beim entkoppelten Verstehen und zum anderen die Prototypisierung der Objekt-Positionierung, die wir über die Kennwerte feststellen konnten. Es ergeben sich mit dem sukzessiven sprachlichen Input die folgenden Aktivierungen:

1. Mit dem Input von *BOS* werden im sprachlichen Output praktisch alle Wörter aktiviert, die satzinitial erscheinen können. Dies schließt nur das Verb *ist*, die Präposition *von* sowie die Satzmarker *EOS* und *BOS* aus – die folgerichtig nicht aktiviert wer-

<sup>32</sup>Wir verwenden folgende Funktionen aus dem Standard-R-Package *stats* (R Development Core Team, 2010): für die ANOVA die Funktion *anova* auf Grundlage eines linearen Modells mit der Funktion *lm*; für den paarweise vergleichenden t-Test die Funktion *pairwise.t.test* mit Korrektur der p-Werte für mehrfache Tests über die Methode nach (Holm, 1979).

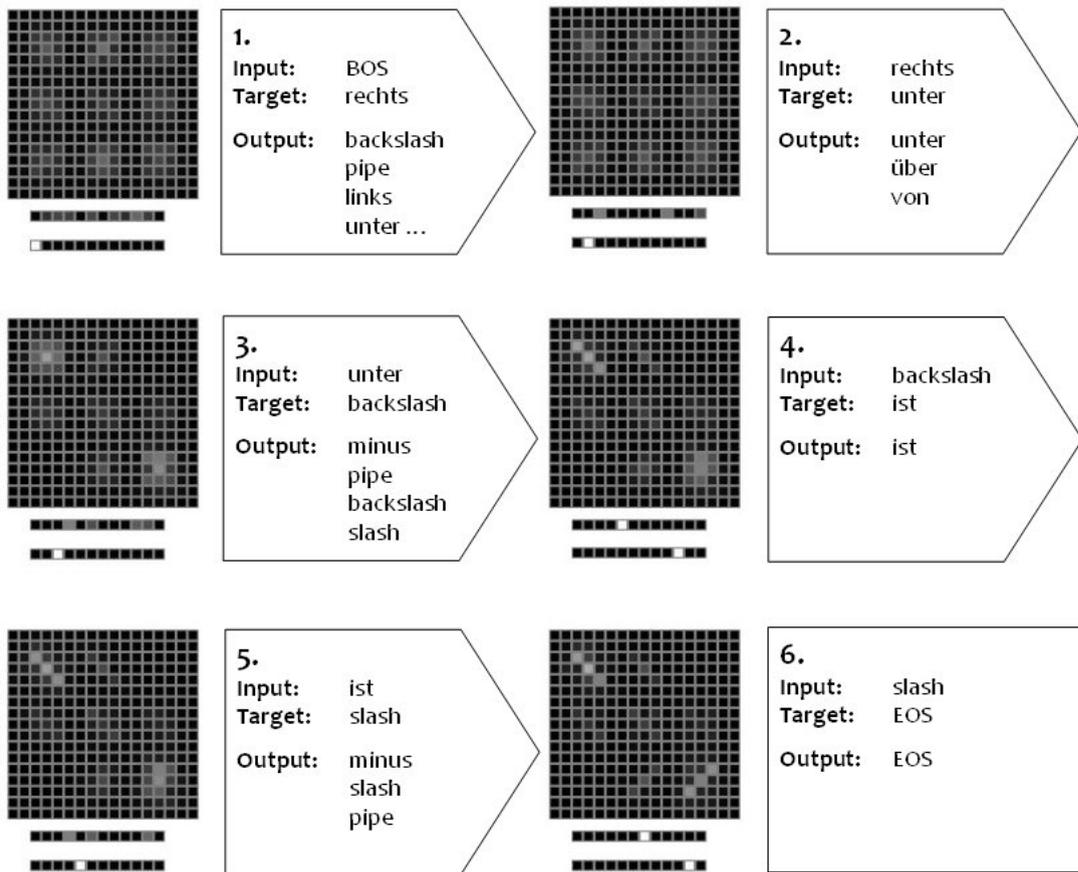


Abbildung 10.26.: SynSem II, Variante B.c: Beispiel zum entkoppelten Verstehen.

den. Im visuellen Output zeigt sich eine diffuse Aktivierung, die sich auf die neun Ideal-Positionen beschränkt.

2. Mit dem Erscheinen des Inputs *rechts* werden im sprachlichen Output ausschließlich die Wörter *unter*, *über* und *von* aktiviert, mit denen die Beschreibung der Relation abgeschlossen werden kann. Im visuellen Output ergibt sich keine nennenswerte Veränderung, es deutet sich im Vergleich zu Satzposition 1 höchstens eine leichte Reduzierung der Aktivierung auf den Positionen an, die für die Relation ÜBER bzw. UNTER besetzt werden würden.
3. Mit dem Wort *unter* zeigt sich im visuellen Output die erste größere Veränderung: Die Aktivierung beschränkt sich weitestgehend auf die Positionen für die Relation RECHTSUNTER. Die leichte Aktivierung der übrigen Positionen ist darauf zurückzuführen, dass diese Relation in Variante B auch zu anderen Positionierungen korrespondieren kann. Einzige Ausnahmen sind dementsprechend die oberste rechte und die unterste linke Position – diese Felder können bei dieser Relation nicht besetzt sein. Wie wir bereits für SynSem I beobachten konnten, werden auf den beiden Hauptpositionen am stärksten die Zentren der Positionen aktiviert. Im sprachlichen Output werden korrekt für diese Satzposition mit etwa gleicher Wahrscheinlichkeit die Nomen für die vier möglichen Objekte aktiviert.
4. Mit dem Input von *backslash* wird korrespondierend die Aktivierung der entsprechenden Position in der Relation RECHTSUNTER angepasst. Im sprachlichen Output wird korrekt ausschließlich das Verb vorhergesagt.
5. Mit dem Input des Verbs *ist* erfolgt im sprachlichen Output die Vorhersage der Nomen für die möglichen Objekte. Ausgenommen ist nur *backslash*, der bereits genannt wurde und dementsprechend nicht mehr erscheinen kann.
6. Mit dem Erscheinen des Nomens *slash* erfolgt schließlich die Konkretisierung des bislang diffus aktivierten Objekts SLASH. Zudem zeigt sich (wie bereits auf den beiden vorhergehenden Satzpositionen für BACKSLASH) eine minimale Aktivierung der Objekte BACKSLASH und SLASH auf Positionen ihres möglichen Erscheinens in der Relation RECHTSUNTER. Im sprachlichen Output wird korrekt *EOS* vorhergesagt. Auf Position 7 ergibt sich außer der Vorhersage für *BOS* keine Veränderung mehr.

**Detailanalyse: Visuelle Route** Welchen Einfluss hat der Unterschied zwischen den drei Trainings-Versionen auf die qualitative Performanz? Abbildung 10.27 zeigt die Ergebnisse für die Objekt-Identifizierung (Abb. 10.27(a)), die Objekt-Lokalisierung (Abb. 10.27(b)) und den Gesamtfehler  $1 - \text{Cos Sem}$  (Abb. 10.27(c)) für die beiden Satztypen ORO und ROO im Vergleich für die drei Versionen B.a (im Diagramm: fcn01), B.b (fcn03) und B.c (fcn05). Für die Objekt-Identifizierung und den Gesamtfehler zeigt sich ein minimaler geringerer Fehlerwert für Version B.a auf den beiden letzten Satzpositionen. Ansonsten zeigt sich ein weitgehend gleicher Verlauf in allen drei Fehlerwerten für die verschiedenen Versionen. Einzige Ausnahme ist der abweichende Fehlerverlauf bei der Objekt-Identifizierung für Version B.b beim Satztyp ROO auf den ersten vier Satzpositionen. Diese schwer erklärable Abweichung wirkt sich jedoch nicht auf den Verlauf des Gesamtfehlers aus. Insgesamt zeigen sich damit für alle drei Versionen die gleichen qualitativen Merkmale für die Konstruktion der visuellen Situation. Für die Objekt-Identifizierung zeigt sich die stärkste

Fehlerminimierung beim Erscheinen des zweiten Nomens. Die Lokalisierung wird hingegen wieder minimal durch das erste und vorwiegend durch das zweite Adverb (bzw. die auf derselben Satzposition erscheinende Präposition) bestimmt.

**Detailanalyse: Sprachliche Route** Für die Wortvorhersage (Abb. 10.28) zeigt sich der Unterschied noch weniger, der Verlauf des Fehlers für die beiden Satztypen ist für die Versionen B.a, B.b und B.c weitgehend deckungsgleich. Es zeigen sich wieder die gleichen Merkmale der Verarbeitung, die wir bereits für Version A beschrieben haben (vgl. Diskussion zu Abb. 10.20).

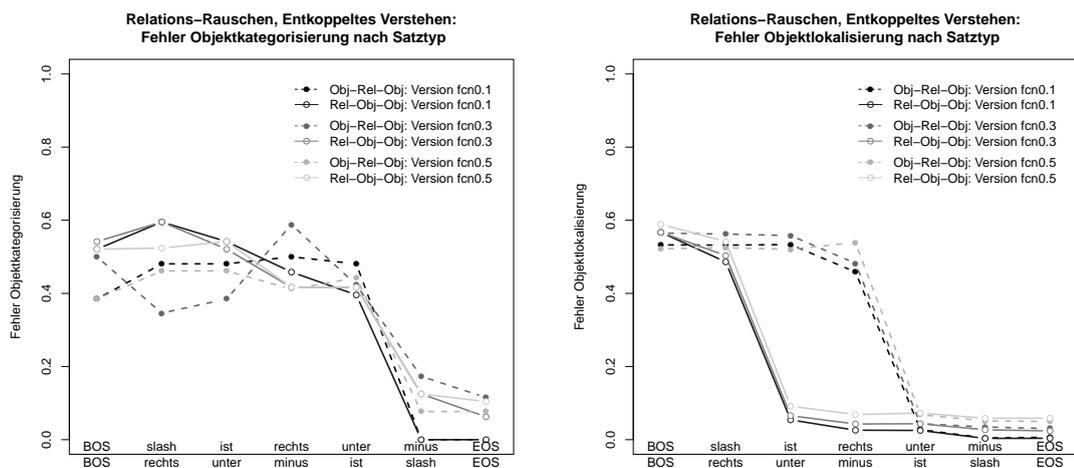
**Zusammenfassung** Wie bereits für Variante A zeigt sich für Variante B eine Prototypisierung der Objektpositionierung, mit der neben einer Generalisierung über Objektverschiebungen durch Positionsrauschen auch eine Generalisierung über größere Objektverschiebungen durch Relationsrauschen demonstriert wird. Der untersuchte Unterschied zwischen den drei Trainingsversionen betrifft den Anteil verschobener Objekt-Konstellationen und damit den Grad, zu dem sich eine Prototypisierung aus der distributionalen Struktur des Inputs ableiten lässt. Die Tatsachen, dass sich qualitativ kein Unterschied zwischen den drei Versionen zeigt und insgesamt betrachtet der etwas erhöhte quantitative Fehler der visuellen Route für Version B.c immer noch recht gering ist, weisen auf eine recht robuste Verarbeitung und eine gute Generalisierungsfähigkeit des Modells hin. Zudem zeigen sich alle Verarbeitungsmerkmale, die wir schon für andere Versionen beschrieben haben.

#### 10.3.4.2. Situierete Sprachverarbeitung (Variante B)

Beim situierten Sprachverstehen spielt es für die visuelle Verarbeitung im Unterschied zum entkoppelten Verstehen keine Rolle, ob auf Idealpositionen oder auf versetzten Relationen getestet wird: Sowohl für die sprachliche als auch für die visuelle Route zeigen sich unter diesen beiden Bedingungen übereinstimmende Werte. Zudem spielt es in diesem Test-Modus auch keine Rolle, welche Trainings-Version zugrundegelegt wird – die Fehlerwerte zeigen eine hohe Übereinstimmung.

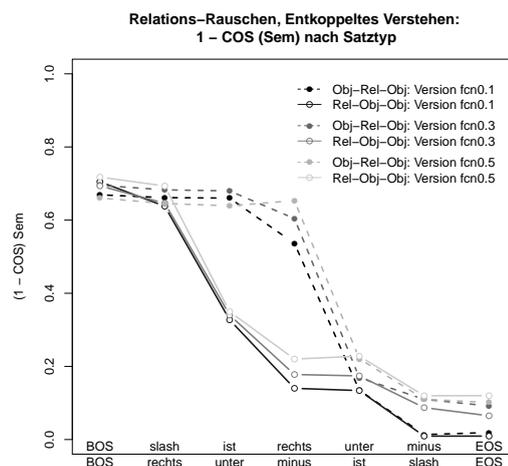
**Globale Kennwerte** Die positionsübergreifenden Fehlerwerte für Tests mit versetzten Positionen, gemittelt über die Versionen mit 10%, 30% und 50% Abweichung, zeigen eine sehr gute Performanz in allen Aspekten mit  $\bar{x}_{\text{LocScore Sem}} = 0.019$  ( $\sigma = 0.049$ ),  $\bar{x}_{1\text{-ObjCor Sem}} = 0.076$  ( $\sigma = 0.19$ ) und  $\bar{x}_{1\text{-Cos Sem}} = 0.049$  ( $\sigma = 0.106$ ). Zugleich ergeben sich aber wie schon in der Variante mit Positionsrauschen keine Verbesserungen durch den sprachlichen Input, denn auf der Satzposition *EOS* halten die Werte mit  $\bar{x}_{\text{LocScore Sem}} = 0.019$  ( $\sigma = 0.046$ ),  $\bar{x}_{1\text{-ObjCor Sem}} = 0.078$  ( $\sigma = 0.195$ ) und  $\bar{x}_{1\text{-Cos Sem}} = 0.047$  ( $\sigma = 0.103$ ) ein gleichbleibendes Niveau.

Für die sprachliche Route zeigt sich das übliche Bild mit einem niedrigen Fehler für die kategoriale Vorhersage und einem im Vergleich zum entkoppelten Verstehen besseren Gesamtfehler:  $\bar{x}_{1\text{-CatCor Syn}} = 0.08$  ( $\sigma = 0.272$ ) und  $\bar{x}_{1\text{-Cos Syn}} = 0.128$  ( $\sigma = 0.293$ ). Für das Test-Set mit ideal positionierten Objekten ergeben sich praktisch identische Werte. Dies demonstriert, dass die gegebene visuelle Information in beiden Fällen gleichermaßen für



(a) Objekt-Identifizierung (1-ObjCor Sem)

(b) Objekt-Lokalisierung (LocScore Sem)



(c) Fehler Objekt-Relation (1-Cos Sem)

Abbildung 10.27.: SynSem II, Variante B: Performanz der visuellen Route beim entkoppelten Verstehen: Objekt-Identifizierung (1-ObjCor), Objekt-Lokalisierung (LocScore) und Gesamtfehler der Abbildung korrekter Objekt-Relationen (1-Cos Sem). Vergleich der Trainings-Versionen B.a (10% Abweichung), B.b (30% Abweichung) und B.c (50% Abweichung).

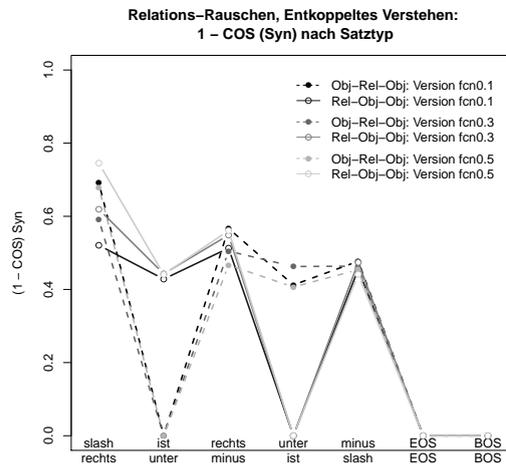


Abbildung 10.28.: SynSem II, Variante B: Ergebnisse für die Wortvorhersage-Performanz beim entkoppelten Verstehen (1-Cos Syn). Vergleich der Trainings-Versionen B.a (10% Abweichung), B.b (30% Abweichung) und B.c (50% Abweichung).

die lexikalische Antizipation genutzt werden kann und sich keine Vorteile aus der idealen Positionierung ergeben.

**Zusammenfassung** Für die Diskussion der situierten Verarbeitung mit Variante B können wir auf Diagramme und Beispiele verzichten. Es zeigen sich für die Fehlerwerte der visuellen Verarbeitung die gleichen Merkmale wie bereits für Variante A (vgl. Abb. 10.23): Es ergibt sich praktisch keine Entwicklung des Fehlers, dieser hält ein gleichbleibend niedriges Niveau. Wie zuvor wird die visuelle Verarbeitung auch in Variante B stark durch den visuellen Input bestimmt, während der sprachliche Input keinen Einfluss nimmt. Die Objekte werden auf den Positionen abgebildet, auf denen sie im Input erscheinen – in Variante B eben mit entsprechend stärker verschobenen Positionierungen. Umgekehrt nimmt der visuelle Kontext aber deutlich Einfluss auf die Verarbeitung der sprachlichen Route: Die Wortvorhersage (1-Cos Syn) wird durch den Kontext verbessert, es zeigt sich wiederum der gleiche Fehlerverlauf, den wir für Variante A mit zwei Input-Objekten erläutert haben (vgl. die Diskussion zu Abb. 10.24).

### 10.3.5. Ergebnisse Variante C: Systematizität

In Variante C haben wir eine verschiedene Anzahl an Relationen für die beiden Objekte SLASH und BACKSLASH vorenthalten und anschließend die Performanz des Netzwerks für genau diese Relationen getestet.<sup>33</sup> In Version C.a erscheint ersteres nie in der Relation RECHTSVON zum zweiten Objekt. In Version C.b ist zusätzlich die Relation ÜBER aus-

<sup>33</sup>Die Positionierung der Objekte im Trainings-Set erfolgt wie in Variante A mit Positionsrauschen mit einer Abweichung um eine Unit in beliebiger Richtung in 15% der Fälle und einer Abweichung um zwei Units in 5% der Fälle.

geschlossen, in Version C.c die Relation RECHTSÜBER und in Version C.d die Relation LINKSÜBER, womit schließlich vier mögliche gemeinsame Relationen der beiden Objekte vorenthalten werden. Das Netzwerk erhält in Version C.d also Input-Situationen, in denen SLASH in den Relationen LINKSVON, UNTER, LINKSUNTER und RECHTSUNTER BACKSLASH erscheint, nicht aber in den komplementären Relationen. Zugleich erscheint SLASH aber in Kombinationen mit den anderen Objekten MINUS und PIPE in den komplementären Relationen. Die Aufgabe des Netzwerks ist es, über diese beiden Erscheinungsformen zu generalisieren – und folglich die nie gesehenen Situationen zu erkennen sowie beim entkoppelten Verstehen einen Satz zu verarbeiten, der zu den vorenthaltenen Situationen korrespondiert. In der Schematisierung für Systematizitäts-Levels nach Niklasson und van Gelder (1994, vgl. Kap. 3.1.4.1) wird damit zumindest Level 1 erreicht: Für dieses muss das Netzwerk erfolgreich Input verarbeiten, bei dem atomare Komponenten zwar auf den Positionen erscheinen, auf denen sie auch im Trainings-Set standen, dabei aber eine neue Formel (d.h. eine neue Kombination der Komponenten) eingehen. Dies ist sicherlich noch keine hinreichende Überprüfung der Systematizität der Verarbeitung unseres Modells, gibt aber einen ersten Eindruck der Generalisierungsfähigkeit und überprüft ein Mindestmaß der Leistungsfähigkeit.

### 10.3.5.1. Entkoppeltes Sprachverstehen (Variante C)

Die unterschiedliche Anzahl der Auslassungen hat sich nicht auf die Performanz ausgewirkt, es konnten in Signifikanztests keine Unterschiede zwischen den Versionen festgestellt werden. Wir stellen die Fehlerwerte im Folgenden gemittelt über alle vier Versionen (C.a, C.b, C.c und C.d) dar. Die Werte wurden für das Test-Set ermittelt, in dem ausschließlich die vorenthaltenen Relationen gegeben sind.

**Globale Kennwerte** Beim entkoppelten Sprachverstehen ergibt sich für die visuelle Verarbeitung der übliche erhöhte Fehlerwert, den wir auch bei den anderen Varianten beobachten konnten:  $\bar{x}_{\text{LocScore Sem}} = 0.263$  ( $\sigma = 0.259$ ),  $\bar{x}_{1\text{-ObjCor Sem}} = 0.368$  ( $\sigma = 0.295$ ),  $\bar{x}_{1\text{-Cos Sem}} = 0.389$  ( $\sigma = 0.3$ ). Die gemittelten Fehlerwerte für die Satzposition EOS zeigen eine deutliche Fehlerreduzierung mit einer Objekt-Lokalisierung von  $\bar{x}_{\text{LocScore Sem}} = 0.039$  ( $\sigma = 0.083$ ), einer Objekt-Identifizierung von  $\bar{x}_{1\text{-ObjCor Sem}} = 0.005$  ( $\sigma = 0.05$ ) und einem Gesamtfehler des visuellen Outputs von  $\bar{x}_{1\text{-Cos Sem}} = 0.101$  ( $\sigma = 0.199$ ).

Für die sprachliche Route zeigt sich das gewohnte Bild eines geringen kategorialen Vorhersagefehlers mit  $\bar{x}_{1\text{-CatCor Syn}} = 0.073$  ( $\sigma = 0.26$ ) und einem erhöhten Gesamtfehler mit  $\bar{x}_{1\text{-Cos Syn}} = 0.337$  ( $\sigma = 0.364$ ). Die geringen Fehlerwerte der visuellen Verarbeitung belegen, dass sich die Auslassungen nicht auf die Performanz der entkoppelten Konstruktion korrespondierender Situations-Abbildungen auswirken.

**Zusammenfassung** Auch für Variante C verzichten wir im Weiteren auf Diagramme und Beispiele. Erstens zeigen sich für die vier verschiedenen Versionen C.a, C.b, C.c und C.d weitgehend übereinstimmende Fehlerverläufe in den unterschiedlichen Kennwerten der visuellen und sprachlichen Verarbeitung. Zudem sind diese weitgehend identisch mit

den Verläufen, die wir für Variante A.a beschrieben haben (vgl. die Diskussionen zu Abb. 10.19 und 10.20).

### 10.3.5.2. Situierete Sprachverarbeitung (Variante C)

**Globale Kennwerte** Für die situierete Verarbeitung ergeben sich positionsübergreifend die geringen Fehlerwerte, die wir schon für Variante B beobachten konnten:  $\bar{x}_{\text{LocScore Sem}} = 0.007$  ( $\sigma = 0.026$ ),  $\bar{x}_{1\text{-ObjCor Sem}} = 0.035$  ( $\sigma = 0.135$ ),  $\bar{x}_{1\text{-Cos Sem}} = 0.022$  ( $\sigma = 0.072$ ). Für die Satzposition nach inhaltlich vollständigem sprachlichem Input ergeben sich praktisch keine Verbesserungen mehr:  $\bar{x}_{\text{LocScore Sem}} = 0.007$  ( $\sigma = 0.025$ ),  $\bar{x}_{1\text{-ObjCor Sem}} = 0.034$  ( $\sigma = 0.131$ ),  $\bar{x}_{1\text{-Cos Sem}} = 0.021$  ( $\sigma = 0.07$ ). Für die Wortvorhersage zeigt sich positionsübergreifend ein niedriger kategorialer Vorhersagefehler mit  $\bar{x}_{1\text{-CatCor Syn}} = 0.072$  ( $\sigma = 0.258$ ) sowie ein sehr geringer Gesamtfehler von  $\bar{x}_{1\text{-Cos Syn}} = 0.097$  ( $\sigma = 0.243$ ). Die recht guten Werte für die visuelle Verarbeitung belegen wiederum, dass die Verarbeitung der ausgelassenen Relationen für das Netzwerk kein Problem darstellt. Der geringe Gesamtfehler für die Wortvorhersage zeigt zudem, dass der neue visuelle Input auch für die lexikalische Antizipation genutzt werden kann.

**Zusammenfassung** Wie zuvor bei der entkoppelten Verarbeitung verzichten wir auch bei der situiereten Verarbeitung für Variante C auf die Abbildung von Diagrammen und Beispielen. Die Fehlerverläufe zeigen die gleichen qualitativen Merkmale, die wir im entsprechenden Test-Modus für Variante A.a beschrieben haben (vgl. Kap. 10.3.3, Abb. 10.23 und 10.24).

### 10.3.6. Ergebnisse zur Sprachproduktion

**Globale Kennwerte: Variante B** Wir ziehen zuerst Variante A mit Positionsrauschen heran. Für die Verarbeitung der visuellen Route zeigen sich positionsübergreifend sehr geringfügige Fehlerwerte:  $\bar{x}_{\text{LocScore Sem}} = 0.012$  ( $\sigma = 0.023$ ),  $\bar{x}_{1\text{-ObjCor Sem}} = 0.065$  ( $\sigma = 0.169$ ),  $\bar{x}_{1\text{-Cos Sem}} = 0.035$  ( $\sigma = 0.068$ ). Dies weist auf die weitgehend sprachunabhängige visuelle Verarbeitung im Modell SynSem II hin. Wenn wir für die Evaluation der Performanz der sprachlichen Route den kategorialen Vorhersage-Fehler verwenden, erhalten wir einen Wert von  $\bar{x}_{1\text{-CatCor Syn}} = 0.288$  ( $\sigma = 0.453$ ), ziehen wir hingegen die Möglichkeit hinzu, beide Satzarten parallel zu vergleichen, so kann bei der kategorialen Vorhersage kein Fehler mehr festgestellt werden ( $\bar{x}_{1\text{-CatCor Syn}} = 0.0$ ). Für die Darstellung des Fehlers im Gesamt-Output der Wortvorhersage-Komponente verwenden wir für die bessere Vergleichbarkeit mit den Ergebnissen aus SynSem I den Fehlerwert EukDist Syn:  $\bar{x}_{\text{EukDist Syn}} = 0.647$  ( $\sigma = 0.579$ ). Während der kategoriale Vorhersagefehler für SynSem II deutlich besser ist als für SynSem I, zeigt sich beim Gesamtfehler nur eine marginale Verbesserung. Beide Fehlerwerte sind allerdings höchst problematisch, wie sich mit der Minimierung des kategorialen Fehlers beim Vergleich der Output-Kategorie mit der Target-Kategorie des jeweils anderen Satzes andeutet: Der hohe Fehlerwert ist teilweise ein Artefakt der Berechnungsgrundlage des jeweiligen Fehlerwerts. Das Modell wählt für die Produktion willkürlich einen Satztyp. Damit ergibt sich vor allem in jenen Beispielen ein hoher Gesamtfehler EukDist Syn, für

die im Test-Set der jeweils andere Satztyp vorgesehen war und für die dementsprechend andere Target-Aktivierungen vorgegeben sind. Zum Teil geht der Fehler also auf eine falsche Vergleichsbasis zurück. Allerdings ist ein Teil des Fehlers auch in falschen Wortfolgen und falschen Beschreibungen begründet.

**Globale Kennwerte: Variante B** Für die Variante B mit Relationsrauschen zeigt sich ein nahezu identisches Bild wie für Variante A: Ein geringer Fehler für die visuelle Verarbeitung mit den Werten  $\bar{x}_{\text{LocScore Sem}} = 0.021$  ( $\sigma = 0.047$ ),  $\bar{x}_{1\text{-ObjCor Sem}} = 0.102$  ( $\sigma = 0.206$ ),  $\bar{x}_{1\text{-Cos Sem}} = 0.058$  ( $\sigma = 0.125$ ), kein Fehler bei der kategorialen Wortvorhersage beim satztypunabhängigen Vergleich mit  $\bar{x}_{1\text{-CatCor Syn}} = 0.0$  und ein erhöhter Fehlerwert für den Gesamt-Output der sprachlichen Verarbeitung mit  $\bar{x}_{\text{EukDist Syn}} = 0.649$  ( $\sigma = 0.596$ ). Für Variante C zeigt sich bei der Systematizitätsprüfung eine noch bessere Performanz. Diese dürfte darin begründet sein, dass das Test-Set ausschließlich Objekt-Konstellationen mit ideal positionierten Objekten enthält: Es ergeben sich für die visuelle Verarbeitung die Fehlerwerte  $\bar{x}_{\text{LocScore Sem}} < 0.001$  ( $\sigma < 0.001$ ),  $\bar{x}_{1\text{-ObjCor Sem}} = 0.0$  (bei satztypunabhängigem Vergleich) und  $\bar{x}_{1\text{-Cos Sem}} < 0.001$  ( $\sigma < 0.001$ ) sowie für die sprachliche Verarbeitung die Fehlerwerte  $\bar{x}_{1\text{-CatCor Syn}} = 0.0$  und  $\bar{x}_{\text{EukDist Syn}} = 0.42$  ( $\sigma = 0.556$ ).

**Beispiel: Beschreibende Sprachproduktion** Zur Demonstration der Sprachproduktion beschränken wir uns auf ein Beispiel für Variante B mit Relationsrauschen: Abbildung 10.26 zeigt die Aktivierungs-Details bei der Produktion des Satzes *slash ist links von minus EOS*.<sup>34</sup> Der zugrundeliegende visuelle Input, der beschrieben werden soll, zeigt die Relation SLASH LINKSVON MINUS in verschobener Positionierung durch Relationsrauschen.

1. Im visuellen Output wird die identische Konstellation und Positionierung der Objekte abgebildet, die im Input vorliegt. An dieser Abbildung ändert sich im Verlauf der Verarbeitung nichts mehr. Die stärkste Output-Aktivierung der sprachlichen Route erhält *slash*.
2. Auf den generierten Input *slash* folgt der sprachliche Output *ist*.
3. Auf *ist* folgt als Output *links*.
4. Auf *links* folgt *von*.
5. Auf *von* folgt *minus*.
6. Auf *minus* folgt *EOS*. Auf der nicht abgebildeten Satzposition 7 wird folgerichtig *BOS* vorhergesagt. Damit wurde nicht nur ein korrekter Satz produziert, sondern auch eine korrekte Beschreibung der visuellen Situation.

Das Beispiel demonstriert die erfolgreiche Beschreibung der gegebenen visuellen Situation mit einem grammatikalisch korrekten Satz. Zudem bezieht sich die Beschreibung auf eine Situation mit stark verschobenen Objekten im Input. Damit wird hier auch für die Produktion gezeigt, dass die Relation der Objekte auch für die verrauchten Situationen korrekt erfasst und für die Verarbeitung der sprachlichen Route genutzt werden kann.

---

<sup>34</sup>Der Pfeil in den Abbildungen erinnert wiederum an den Feedback-Mechanismus mit Winner-takes-all-Funktion für den sprachlichen Output und die Eins-zu-eins-Kopie des Outputs zum Input der sprachlichen Route.

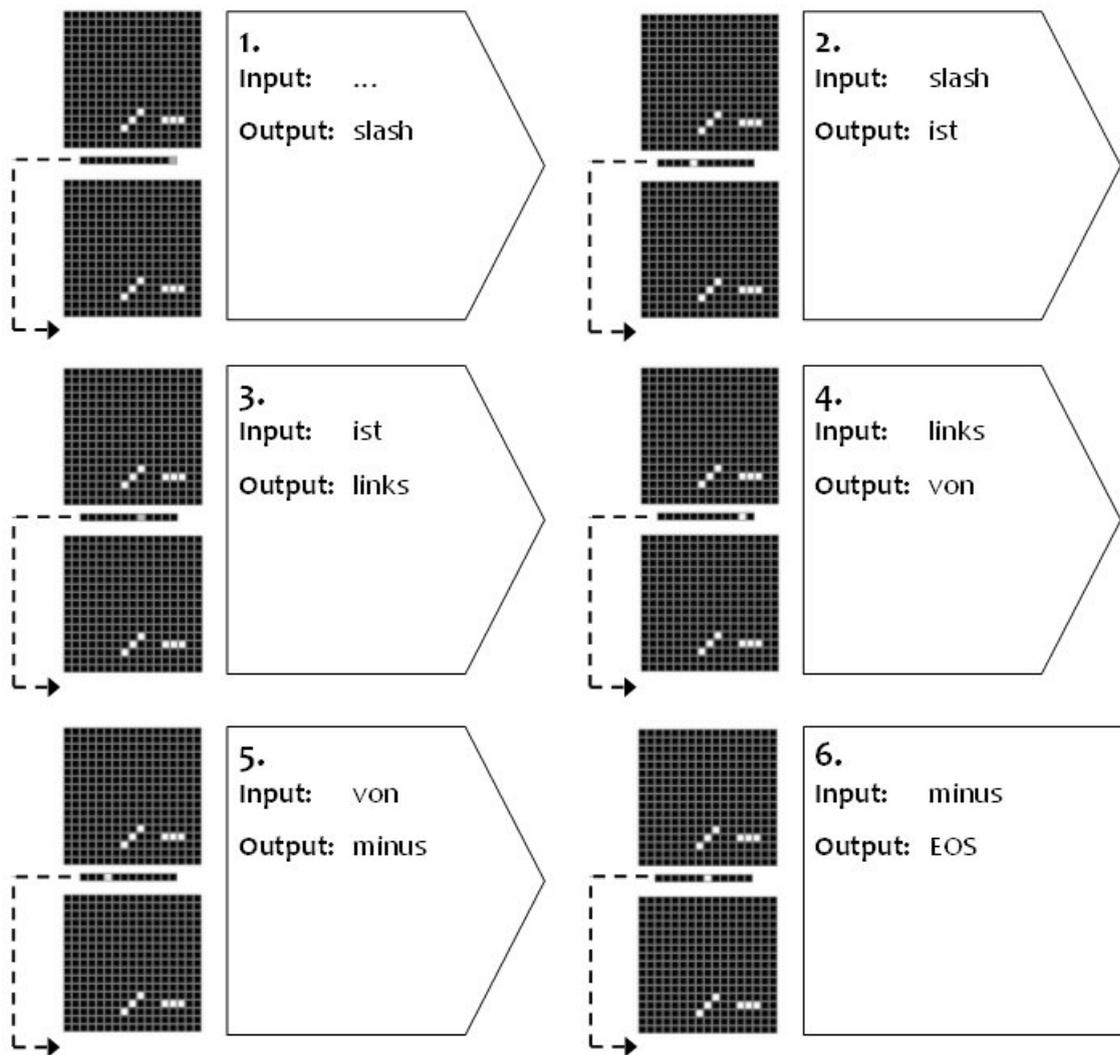


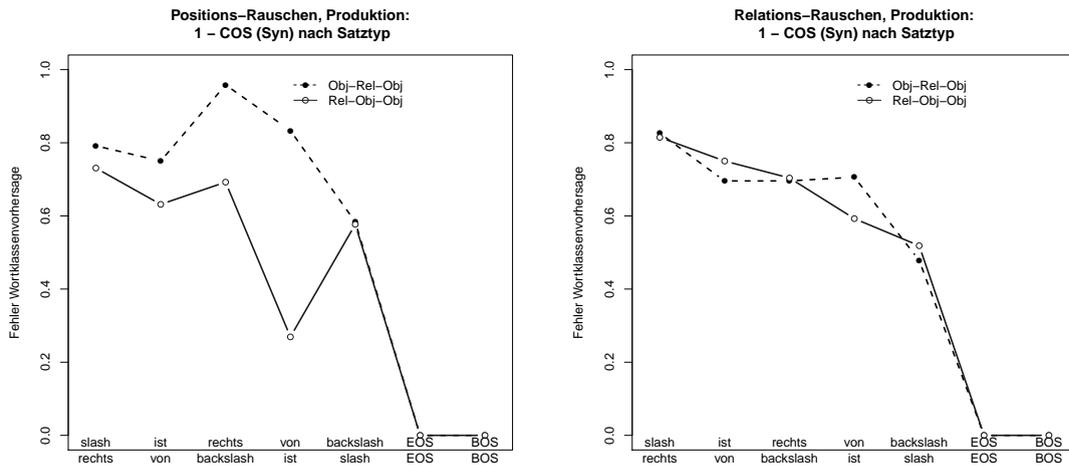
Abbildung 10.29.: SynSem II, Variante B.c: Beispiel für die beschreibende Sprachproduktion für eine Situation mit zwei Input-Objekten.

**Detailanalyse: Sprachliche Route** Die folgenden Diagramme in Abbildung 10.30 stellen die Ergebnisse des Gesamtfehlers 1-Cos Syn der Sprachproduktion in SynSem II für die drei unterschiedlichen Varianten dar. In den Varianten A und B zeigt sich jeweils erst auf der Satzposition *EOS* eine Minimierung des Fehlers, auf den früheren Satzpositionen bleibt er eher hoch. Während die Fehlerverläufe eine eher mäßige Produktionsleistung suggerieren, zeigte sich jedoch bei der Sichtung von Hand, dass überwiegend korrekte Sätze und auch korrekte Beschreibungen produziert wurden. Die häufigsten Fehler stammen von Wiederholungen des bereits genannten Objekts als zweitgenanntes Objekt. Allerdings konnten wir bislang keine geeignete Auswertungsmethode implementieren, die die Überprüfbarkeit der syntaktischen und semantischen Korrektheit des gesamten Satzes gewährleistet. Besonders deutlich wird dies für den Fehler für Variante C (Abb. 10.30): Während sich der Fehler für den Satztyp ORO erst auf der Satzposition 5 minimiert, ist er für den Satztyp ROO bereits von Beginn an minimal. Ein genauerer Blick in die Daten zeigt, dass durch das Netzwerk als Beschreibung immer den Satztyp ROO mit initialem Adverb gewählt wurde. Der vermeintliche Fehler für den Satztyp ORO resultiert also aus einem Vergleich mit den Target-Aktivierungen eines falschen Satztyps.

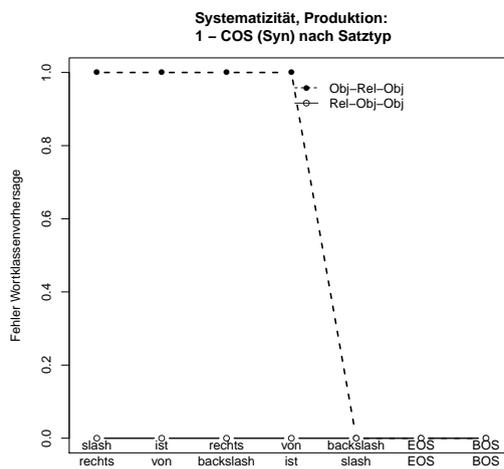
**Zusammenfassung** Im Unterschied zur Sprachproduktion in Simulation SynSem I (vgl. Kap. 10.2.3.3) erfolgt die Sprachproduktion in SynSem II sehr erfolgreich. Das Hauptproblem der Analyse liegt in der vergleichsweise komplexen Auswertung der Daten, da in unserem Vorgehen keine Vorgaben für die Satzstruktur gemacht wurden. Dies hat sich bereits für die Satzproduktion bei SynSem I als Problem herausgestellt – dort allerdings eher bezüglich der mangelhaften Produktionsleistung. Für weitere Simulationen wäre es dementsprechend empfehlenswert, für die ersten ein oder zwei Satzpositionen einen initialen Input zu geben, über den vorgegeben wird, welcher Satz produziert werden soll. Hierdurch dürfte nicht nur die Variabilität der produzierten Sätze erhöht werden, auch die Kontrollierbarkeit der semantischen und syntaktischen Konsistenz des jeweils produzierten Satzes und seine Korrespondenz zur beschriebenen Situation wären auf einfachere Art und Weise gewährleistet.

### 10.3.7. Diskussion von SynSem 2

Das Ziel der Simulation SynSem II war es, in einem zu Simulation SynSem I vergleichbaren Leistungsspektrum den situierten Erwerb und die Verarbeitung von Sprache auf Grundlage einer gänzlich selbstorganisierten visuellen Verarbeitung ohne Vorgabe spezifischer Zielstrukturen zu modellieren. Wir haben hierfür eine auto-assoziative Alternative implementiert, in der die Eigenschaften, die wir in SynSem I explizit enkodiert haben, auf die distributionalen Eigenschaften der Stimuli übertragen wurden. Die unterscheidbaren Objekte erscheinen mit verschiedener Wahrscheinlichkeit auf verschiedenen Positionen des Unit-Gitters des visuellen Inputs und Outputs. Dabei haben wir Varianten mit Positionsrauschen (Verschiebungen der Objekte auf Positionen innerhalb idealer Felder) und Varianten mit Relationsrauschen (zusätzliche Verschiebungen der Objekte über ideale Felder hinweg) erstellt. Eine einfache Erweiterung des Modells um rezeptive Felder sollte die Performanz durch die Herstellung echter Nachbarschaftsbeziehungen im visuellen Grid unterstützen. Die Ergebnisse von SynSem II decken sich teilweise mit den Resultaten und Eigenschaften,



(a) Fehler Wortvorhersage (1-Cos Syn), Variante A (b) Fehler Wortvorhersage (1-Cos Syn), Variante B



(c) Fehler Wortvorhersage (1-Cos Syn), Variante C

Abbildung 10.30.: SynSem II: Performanz der sprachlichen Route bei der beschreibenden Sprachproduktion im Vergleich für die drei Varianten A, B und C.

die wir in SynSem I beobachten konnten. Einige Eigenschaften weichen jedoch deutlich von denen der ersten Simulation ab. Wir fassen kurz die Ergebnisse zusammen, leiten eine Reihe von Eigenschaften daraus ab und ziehen weitergehende Schlußfolgerungen daraus.

Eine erste Eigenschaft betrifft die visuelle Verarbeitung beim situierten Sprachverstehen. In SynSem II wird die visuelle Verarbeitung durch den visuellen Input determiniert. Die Objekte werden unabhängig von der Art des Rauschens auf genau den Positionen erfasst, auf denen sie im Input erscheinen. Dieses Ergebnis unterscheidet sich zu den Eigenschaften in SynSem I. Allerdings wäre zu hinterfragen, ob dieses Verhalten nicht sogar wünschenswert ist. Zum einen leisten Kleinkinder die Kategorisierung von Objekten und die Differenzierung räumlich-relationaler Zusammenhänge schon vor dem Erwerb korrespondierender sprachlicher Begriffe.<sup>35</sup> Zum anderen scheint es intuitiv als realistischere Annahme, dass wir bei der Wahrnehmung und bei der situierten Sprachverarbeitung Objekte nicht nur in relativer, sondern auch in absoluter Positionierung erfassen. Wir können abseits der relativen Positionierung von Objekten (etwa: ‚links über‘) gleichfalls die Objektpositionierung in feinerer Granularität erkennen (etwa: ‚knapp links darüber mit stärkerer Verschiebung nach links‘), die eine beliebige andere Kategorisierung der Relationen erlaubt. Dies könnte in SynSem I kaum umgesetzt werden, ohne erneut explizite Vorgaben zu enkodieren. Die geringere Abhängigkeit von sprachlichen Informationen in SynSem II und die Unabhängigkeit der Abbildung von vorgegebenen Idealisierungen bei der situierten Verarbeitung hingegen erleichtert die Erweiterung des Modells um weitere Kategorisierungen mit feiner Granularität. Dass die Objekte tatsächlich kategorisiert und ihre Relationen differenziert werden, zeigt sich beim entkoppelten Sprachverstehen anhand der entsprechenden Eigenschaften.

Eine zweite Eigenschaft der visuellen Verarbeitung beim situierten Sprachverstehen betrifft die Reduzierung der Objekte. Zunächst einmal erfolgt eine Reduzierung nur dann, wenn das Training mit einer reduzierten Anzahl von Objekten erfolgt. Werden im Training drei Objekte verwendet, zeigt sich keine Einschränkung berücksichtigter Objekte. Beim Training mit zwei Objekten erfolgt die Reduzierung wiederum über den visuellen Input, für den wir ideale Positionierungen mit gegenüberliegenden Positionen als Kriterium in den Aufbau der Stimuli übertragen haben. Zwar ergibt sich dadurch eine automatische Reduzierung der korrespondierenden abgebildeten Situation auf die relevanten Aspekte der Input-Situation. Durch die Vorgabe der Reduzierung über die Struktur des visuellen Inputs erfolgt diese Eigenschaft aber wiederum unabhängig vom sprachlichen Input. Dadurch verlieren wir mit SynSem II leider die Inferenzmechanismen im Sinne einer Lenkung der visuellen Aufmerksamkeit. Zudem erscheint die Abhängigkeit der Reduzierung von den Merkmalen des visuellen Inputs unrealistisch. Dies würde bedeuten, dass wir keine beliebigen Relationen zwischen beliebigen Objekten erfassen könnten. Wir haben allerdings nicht untersucht, welche Auswirkung Input-Situationen mit drei Objekten auf die Verarbeitung in Variante B haben. Durch das Relationsrauschen können auch Relationen zwischen nicht ideal positionierten Objekten erfasst werden. Damit wird die starre künstliche Vorgabe wahrnehmbarer Relationen aufgehoben. Wir gehen davon aus, dass sich unter diesen Voraussetzungen für die Darbietung von drei Objekten wieder die Effekte der Lenkung der visuellen Aufmerksamkeit ergeben, die wir in Simulation SynSem I beobachten konnten.

---

<sup>35</sup>Vielen Dank für diesen Hinweis an einen anonymen Reviewer eines Artikels zu dieser Studie.

Eine dritte Eigenschaft betrifft die sprachliche Verarbeitung beim situierten Verstehen. Hier zeigen sich die Eigenschaften der durch den situationalen Kontext gestützten Antizipation von Folgewörtern, die wir für SynSem I beobachten konnten. Die Antizipation bezieht sich dabei nicht nur auf einzelne Aspekte der Situation, sondern auch auf ihren gesamten Zusammenhang – denn die Folgewörter werden auch in Beziehung zur jeweils möglichen räumlichen Relation bestimmt. Der Einfluss zeigt sich wiederum am deutlichsten beim Vergleich mit der Performanz der sprachlichen Route bei der entkoppelten Verarbeitung, bei der die Wortvorhersage nur aufgrund von Kollokationsinformationen erfolgen kann, während inhaltliche Aspekte nicht berücksichtigt werden.

Eine vierte Eigenschaft betrifft die sprachliche Indizierung von Aspekten visueller Situationen und das darauf beruhende Re-Enactment bei der Konstruktion entsprechender Situationen durch Sprache beim entkoppelten Verstehen. Auch für diesen Aspekt zeigt sich die Bedeutung der künstlich vorgegebenen Reduzierung der Objekte im Training: Die Performanz beim entkoppelten Verstehen für die Version, die mit drei Objekten trainiert wurde (Version A.b), zeigte deutliche Anzeichen, dass die Indizierung nur sehr rudimentär erworben wurde. In den Versionen, die mit nur zwei Objekten trainiert wurden (Version A.a sowie die Varianten B und C), wurde die Indizierung hingegen sehr erfolgreich erworben. Dies zeigte sich zum einen in der wesentlich besseren Performanz und zum anderen in der Sensitivität der Konstruktion spezifischer Situationsaspekte für bestimmte Wortklassen (die Sensitivität der Objekt-Identifizierung von Nomen und der Objekt-Lokalisierung von Adverbien und Präpositionen). Umgekehrt können wir aus der erfolgreichen Indizierung und dem erfolgreichen Re-Enactment für die entsprechenden Versionen folgende Annahme ableiten: Die Reduzierung der Objektmenge ist unmittelbar relevant für den Bedeutungserwerb und muss durch einen kognitiven Mechanismus geleistet werden. In dieser Hinsicht ist auch die Konstruktion der Stimuli in SynSem II wieder eine künstliche Vorgabe, denn für die Trainingsphase geben wir gewissenmaßen eine entsprechende Reduzierung auf zwei Objekte vor, indem wir immer nur zwei Objekte darbieten. Ist diese Grundlage einmal geschaffen, können auch einzelne Relationen aus Situationen mit mehr als zwei Objekten erfasst werden. Wir werden in der Abschluss-Diskussion nochmal darauf zurückkommen, wie eine solche Reduzierung kognitiv über eine beschränkte Wahrnehmungskapazität und durch geteilte und geleitete Aufmerksamkeit im Erwerb realisiert sein könnte.

Eine fünfte Eigenschaft betrifft die prototypisierte Positionierung beim entkoppelten Sprachverstehen. Anders als in SynSem I haben wir diese indirekt über die distributionalen Eigenschaften des Inputs induziert. Und anders als in SynSem I wirkt sich diese Eigenschaft in SynSem II nur bei der entkoppelten Verarbeitung aus. Durch die hier erfolgende Prototypisierung zeigt sich wie oben bereits erwähnt zum einen, dass die Objekte des visuellen Inputs tatsächlich unterschieden und dass die möglichen Relationen der Objekte differenziert und kategorisiert werden. Allerdings werden auch hier die Kategorisierungsprozesse über die Sprache vermittelt. Wir können aus unseren Simulationen nicht ableiten, ob die Kategorien auch sprachunabhängig unterschieden werden. Beim entkoppelten Verstehen zeigt sich auch wieder die Inkrementalität der Verarbeitung, die sich im Modell in der unmittelbaren Konstruktion möglichst vollständiger Objekt-Relationen und unterspezifizierter Objekt-Repräsentationen äußert.

Die sechste Eigenschaft schließlich betrifft die Sprachproduktion, die in SynSem II deutlich erfolgreicher simuliert werden konnte – auch wenn wir dies mangels ausgereifter

Analyseverfahren nicht in letzter Konsequenz demonstrieren konnten. Allerdings konnten wir bislang keinen Grund identifizieren, der für die bessere Qualität der Performanz verantwortlich ist. Wie bereits für beide Simulationen angemerkt, fehlt bei dieser Form der Modellierung von Sprachproduktion allerdings ein Mechanismus, über den Strukturvorgaben getroffen werden. Zudem wirkt sich die fehlende Kontrolle der semantischen und syntaktischen Konsistenz im Modell direkt auf die Qualität der Performanz aus, in der sich typische Fehlleistungen wie zum Beispiel Wortwiederholungen finden. Mit Blick auf den Gegenstand der Modellierung scheint uns die Annahme eines gänzlich unkontrollierten, willkürlich agierenden Sprachproduktionsmechanismus, der nicht über Kontrollmechanismen für die Konsistenz der produzierten Äußerung verfügt, kein realistisches Modell menschlicher Sprachproduktion zu sein.

## 10.4. Perspektive von SynSem auf Sprachverarbeitung

Im dritten Teil dieser Arbeit haben wir die Perspektive der Embodied Cognition auf die Sprachverarbeitung dargelegt. Wir haben untersucht, inwiefern sich die Konzepte des Embodied Language Processing mit Konzepten des Konnektionismus decken und ob ein wechselseitig ergänzender Ansatz eine fruchtbare Perspektive und Modellierungsgrundlage für Sprachverstehensprozesse eröffnet. Im dritten Kapitel des dritten Teils haben wir nun zwei Simulationsserien vorgestellt, in denen wir verschiedene Aspekte des Sprachverstehens im Sinne des Language Grounding und des Embodied Language Processing modelliert haben. Dabei haben wir die folgenden Gegenstände behandelt: (a) den situational verankerten Bedeutungserwerb, (b) die Etablierung von Eins-zu-eins-Korrespondenzen bei potenzieller Viele-zu-viele-Korrespondenz zwischen Mehrwort-Äußerungen und komplexen visuellen Szenen, (c) die Indizierung von nicht-sprachlichen Inhalten durch sprachliche Formen, (d) die Emergenz von Sprachverarbeitungsfähigkeiten aus der Integration verschiedener Informationsstrukturen und Transformationsaufgaben, (e) das situational entkoppelte Sprachverstehen als Re-Enactment von Derivaten modaler Erfahrungen, (f) die Sprachproduktion als situationale Beschreibung durch das Re-Enactment sprachlicher Formen und (g) Inferenzen im Sinne von sprachlich gelenkter Aufmerksamkeit und modalitätsübergreifendem Priming bei der situierten Sprachverarbeitung.

Abschließend diskutieren wir, inwiefern wir mit unserem Modellierungsansatz die Kriterien erfüllen konnten, die wir in Kapitel 9.3.3 postuliert haben. Wir bewerten unsere Modellierungsergebnisse in Bezug auf den Modellierungsgegenstand Spracherwerb und -verarbeitung und schlagen Erweiterungsmöglichkeiten für zukünftige Modellierungen in diesem Ansatz vor. Wir skizzieren die Perspektive auf Sprachverstehensprozesse, die wir mit diesem Ansatz vorschlagen, und besprechen Konsequenzen und Anregungen für Theorien des Embodied Language Processing.

### 10.4.1. Modellierungskriterien

Als Zielvorgabe für die Modellierung von Spracherwerb und -verarbeitung im Sinne des Embodied Language Processing haben wir die Kriterien Analogizität, Multimodalität, Informationsintegration, Situationalität, Emergenz und Generalisierung formuliert. Erfüllen

unsere Simulationen SynSem I und II diese Kriterien? Unserer Einschätzung nach können wir diese Frage mit einem Ja mit Einschränkungen beantworten: Unser Modell ist in Hinsicht auf unsere Vorgaben sicherlich nicht perfekt, im Unterschied zu anderen Modellen syntaktisch-referenzieller Verarbeitung, die in Beziehung zur Embodied Cognition gestellt wurden, erreicht es jedoch eine wesentlich höhere Übereinstimmung.

- Die Analogizität ist gewährleistet, denn die relevanten Teilausschnitte des Inputs werden erfasst und in die gleiche oder zumindest eine ähnliche Struktur überführt, ohne eine Änderung des Formats vorzunehmen. Allerdings wurden bei SynSem I durch die Vorgabe der Prototypen in den Target-Vektoren explizit Informationsstrukturen vorgegeben, die in dieser Form nicht im Input gegeben waren. Bei SynSem II erfolgte eine ähnliche Vorgabe durch die Reduzierung der Objektanzahl im Training.
- Die Modalität wurde durch die Enkodierung der nicht-sprachlichen Inhalte als visuelle Stimuli erfüllt, die sich wie vorgegeben an einer Modalität der natürlichen kognitiven Verarbeitung orientiert. Allerdings gibt es auch hier Einschränkungen: Konsequenter wäre es gewesen, wenn auch die sprachlichen Stimuli nach dem Vorbild akustischer Signale gestaltet worden wären. Zudem haben wir durch die Gestaltung der Stimuli gewisse Verarbeitungseigenschaften begünstigt.
- Die Informationsintegration im Sinne multimodaler Assoziation wurde durch die parallel-interaktive Architektur und das holistische Training gewährleistet. Dass tatsächlich eine Integration stattfand und nicht nur eine gemeinsame, aber unabhängige Nutzung der gemeinsamen Ressourcen, wurde durch die Ergebnisse belegt.
- Durch die holistische Verarbeitung, die sich von anderen Modellierungen mit getrennten Trainingsphasen für unterschiedliche Module (vgl. Kap. 9.3.2) unterscheidet, erfolgt der Spracherwerb situiert und auf den situationalen Kontext bezogen.
- Die relevanten Verarbeitungseigenschaften haben sich emergent aus der Integration der Informationsstrukturen und aus der Interaktion der Aufgaben entwickelt. Wiederum haben wir aber bestimmte Aspekte der sich entwickelnden Eigenschaften begünstigt – etwa durch die Vorgabe der prototypisierten Positionierungen in SynSem I oder durch die Reduzierung der Objekte im Training in SynSem II.
- Über die Performanz der Netzwerke konnten wir verschiedene Generalisierungsleistungen belegen. Die Leistungen wurden für Materialien demonstriert, die unabhängig von Trainings-Set generiert wurden. Zudem wurde für SynSem II eine niedrigere Stufe der Systematizität belegt.

### 10.4.2. Eigenschaften des Modells

**Betreffend expliziter Modellvorgaben** Ein Hauptproblem unseres Ansatzes besteht sicherlich darin, wie wir in unserem Modell die Reduzierung und Generalisierung visueller Informationen implementieren. Die Simulationen in SynSem II waren nur dann erfolgreich, wenn wir ein Training mit nur zwei Input-Objekten vorgegeben haben. In SynSem I wurde eine entsprechende Reduzierung im Training durch eine Vorgabe der auszuwählenden Objekte spezifiziert. Dabei wurde sogar explizit eine Prototypisierung der relationalen Positionierung vorgegeben. Diese wurde in SynSem II implizit durch die distributionale

Struktur der Stimuli enkodiert. Damit muss das Netzwerk zumindest eine Ableitung der relevanten Eigenschaften des Inputs leisten. Dennoch bleibt zu fragen, ob unsere distributionalen Vorgaben nicht ebenfalls einer Form von versteckt expliziter Strukturenkodierung entsprechen. Jedenfalls zeigt die Erfordernis einer gezielten Gestaltung der Stimuli, dass wir die Verarbeitung bislang immer noch in einem gewissen Maß leiten oder strukturieren müssen. Akzeptieren wir die Vorgaben zunächst einmal als sinnvolle und realistische Annahmen, können wir feststellen, dass die kategorial-schematische Wahrnehmung, die Reduzierung und die Prototypisierung wichtige Faktoren für die zu erklärenden kognitiven Leistungen sind. Durch sie kann das Modell situierten Bedeutungserwerb und die komplexe Abbildung von Form auf Inhalt gewährleisten. Unser kognitives System und die Interaktion mit der Umwelt müssen Mechanismen und Prinzipien aufweisen, über die die Wirkung dieser Faktoren gewährleistet wird.

Wir haben in Kapitel 9.1.3 verschiedene Ansätze besprochen, in denen das Problem des Vokabelerwerbs und der Etablierung von Eins-zu-eins-Korrespondenzen für die Abbildung von Form auf Bedeutung gewährleistet. In diesen Ansätzen wurden verschiedene Formen des (statistical) cross-situational learning (z.B. Yu & Smith, 2006) beschrieben. Auch in unserem Modell wird der komplexe Bedeutungserwerb durch die statistische Erfassung von Übereinstimmungen zwischen sprachlicher Form und situationaler Bedeutung über viele Situationen hinweg erfasst. Als Grundlage wurden in den verschiedenen anderen Modellen Mechanismen der Auswahl (Reduzierung) und der Generalisierung (Prototypisierung) implementiert, um die Voraussetzungen für den situationsübergreifenden Lernprozess zu schaffen (vgl. u.a. Siskind, 1996; Roy, 1999; Davis, 2002; Yu & Ballard, 2003, 2004; Yu, 2006). Während der statistische Lernprozess und die Etablierung korrespondierender Form-Bedeutungs-Paare in unserem Modell durch inhärente konnektionistische Mechanismen gewährleistet wird, bedürfen die über die Stimuli realisierten Begünstigungen der Reduzierungs- und Prototypisierungsprozesse einer Begründung.

Folgende Mechanismen und Interaktionsprinzipien könnten unser Vorgehen rechtfertigen: (a) Eigenschaften der Stimuli wie etwa Salienz, durch die in der frühen Erwerbsphase nur eine beschränkte Menge an Objekten und Situationsaspekten verfolgt wird; (b) Eigenschaften der sozialen Interaktion, durch die eine gemeinsame gerichtete Aufmerksamkeit hergestellt und eine nicht-sprachlich induzierte Lenkung der Aufmerksamkeit durch den Sprecher geleistet werden; (c) physische Eigenschaften wie etwa Unterschiede der Sehschärfe des Auges in verschiedenen Bereichen des Gesichtsfelds mit zentralem (auf der Fovea centralis) und peripherem Sehen, durch die der situationale Ausschnitt der Welt über die physiologischen Bedingungen reduziert wird; (d) Filterung durch die modalitätenübergreifende Nutzung von Vorwissen im Sinne eines wechselseitigen Bootstrapping-Mechanismus. Erstrebenswert wäre es, wenn das Modell solche Prinzipien inhärent und ohne äußere Vorgaben umsetzen würde. Bis diese Aspekte in das Modell integriert sind, müssen wir uns jedoch damit zufrieden geben, dass es für die Vorgaben unserer Stimuli zumindest potenzielle Erklärungen und damit für unsere Vorgehensweisen Rechtfertigungen gibt.

**Spracherwerb und -verarbeitung** Die Eigenschaften der Sprachverarbeitung, des situierten Spracherwerbs und des Verstehensprozesses wurden bereits in den Diskussionen innerhalb der Unterkapitel zu den beiden Modellen SynSem I und SynSem II erläutert (vgl.

Kap. 10.2.4 und 10.3.7). In den Diskussionen finden sich neben den Leistungsmerkmalen der Modelle auch Besprechungen der Vergleichbarkeit dieser Merkmale mit Befunden zu korrespondierenden Gegenständen aus der Psycholinguistik – wofür vor allem Befunde zur engen Beziehung zwischen sprachlicher und visueller Verarbeitung relevant werden (u.a. Tanenhaus et al., 1995; Altmann & Kamide, 1999; Stanfield & Zwaan, 2001; Zwaan et al., 2002; Dahan & Tanenhaus, 2005; Knoeferle & Crocker, 2006; Altmann & Kamide, 2007, 2009). Zudem haben wir Bezüge zu Ansätzen des Embodied Language Processing erläutert – situationale Verankerung im Erwerb, sprachliche Indizierung modaler Repräsentationen (Glenberg & Robertson, 2000), Verstehen als Re-Enactment von Derivaten früherer Erfahrungen (Barsalou, 1999; Barsalou et al., 2003) und als Simulation der Wahrnehmung korrespondierender Situationen (Zwaan, 2004).

Fassen wir nochmals kurz zusammen, welche Eigenschaften unser Modell nahelegt: Sprachverstehen erfolgt als inkrementeller Prozess der Konstruktion eines Situationsmodells, das so ausführlich wie mit der aktuell gegebenen Information möglich vervollständigt wird. Unvollständige Informationen werden über unterspezifizierte Repräsentationen der Sachverhalte erfasst. Das Modell wird bei neuen verfügbaren Informationen unmittelbar aktualisiert, wobei gleichermaßen sprachliche als auch bei der situierten Verarbeitung visuelle Informationen genutzt werden. Dabei zeigt sich eine starke wechselseitige Korrespondenz zwischen sprachlicher und visueller Verarbeitung mit einer direkten domänenübergreifenden Nutzung der Informationen, die sich als gegenseitige Constraints für die sprachliche Antizipation und die visuelle Aufmerksamkeit äußern und als Quelle für Inferenzen zu möglichen situationalen Interpretationen dienen. Dabei zeigen die Befunde für das Modell SynSem I, wie die Lenkung der visuellen Aufmerksamkeit im Sinne der Annahme von Altmann und Mirkovic (2009) und entgegen der expliziten Implementierung von Mayberry et al. (2009) als emergentes Produkt der Aufgaben-Koordination und -Integration implizit im Modell erfolgen kann.

### 10.4.3. Bedeutung im SynSem-Ansatz

Wir haben im Verlauf dieser Arbeit verschiedene Konzeptionen von Bedeutung im Konnektionismus kennengelernt (vgl. u.a. die Diskussionen in Kap. 3.1.3.2, 4.5 und 9.3). Welche Konzeption von Bedeutung wird mit dem Modell SynSem etabliert, und wo wird sie wie repräsentiert? Eine zentrale Stellung nimmt zunächst einmal der Hidden-Layer ein, über den die Koordination der Aufgaben und die domänenübergreifende Integration von Informationen vermittelt wird. In anderen Ansätzen mit Dual-Route-Architektur wird auf die zentrale Bedeutung des Hidden-Layers hingewiesen, der als Kern des semantischen Systems dient und dessen Repräsentationen als Substrat der Bedeutung verstanden werden (vgl. Plunkett et al., 1992; Cangelosi et al., 2002). Über ihn werden die wechselseitigen Beziehungen und gegenseitigen Constraints verschiedener Informationssysteme ausgehandelt. Übertragen auf den Ansatz des Embodied Language Processing repräsentiert er den Mechanismus, über den die sprachliche Indizierung und das sprachliche Re-Enactment bei der Simulation eines Situationsmodells im Verstehensprozess etabliert werden. Zudem wird über die repräsentationale Kapazität des Hidden-Layers, über seine Möglichkeit der Ad-hoc-Generierung neuer Repräsentationen sowie über seine kontextabhängige Modulier-

nung von Repräsentationen das semantische Potenzial des Modells gewährleistet (vgl. auch Cangelosi, 2005).

Allerdings nehmen wir nicht an, dass sich das konzeptuelle Wissen in unserem Modell auf die Repräsentationen des Hidden-Layers beschränken oder dass sich dieses Wissen als einheitliche Informationsstruktur charakterisieren lässt. Zum einen werden die Repräsentationen über die Gewichtsmatrizen aus den verschiedenen Domänen bestimmt, die die modalitätsspezifischen Constraints bestimmen. Zum anderen enthalten die transportierten Informationen jeweils Anteile, die nur innerhalb einer modalen Route bzw. für deren Aufgabe relevant sind. Dies entspricht der Annahme einer graduell funktionellen modalitätsspezifischen Spezialisierung (vgl. Plaut, 2002; Rogers & Plaut, 2002). Sprachliche Bedeutung könnte in diesem Modell Elmans operationaler Auffassung folgend als Erzeugung spezifischer Systemzustände beschrieben werden, wobei die Bedeutung von Wörtern den kausalen Effekten der Wörter auf die mentalen Zustände entspricht (vgl. Elman, 2004, S. 306).

Über die konnektionistischen Aspekte der Erfassung von Bedeutung hinaus ist für unser Modell die Erfassung der Perspektive der Embodied Cognition relevant. Als zentraler Aspekt wird etwa im Ansatz von Zwaan (2004) der Einfluss der Eigenschaften unserer Wahrnehmungsmodalitäten auf das Sprachverstehen hervorgehoben: Semantik ist nichts rein Sprachliches – sie ist zentral davon beeinflusst, wie wir die sprachlich benannten Referenten wahrnehmen bzw. wahrgenommen haben (prototypisierte und schematisierte Derivate unserer Erfahrungen in der Interaktion mit der Umwelt). Das bedeutet, dass bei einer Theorie sprachlicher Semantik die multimodalen Eigenschaften der Wahrnehmungsderivate und der motorischen Affordanzen sowie deren Erwerb berücksichtigt werden müssen. Diese Idee geht insofern in unser Modell ein, als dass sprachliche Information direkt mit visuellen Informationen und deren Eigenschaften assoziiert sind. Beim entkoppelten Verstehen führt genau diese Assoziation mit den zentralen Tendenzen aller Situationen, die mit den Elementen eines Satzes aufgetreten sind, zur Konstruktion eines prototypischen Situationsmodells. Die Abhängigkeit sprachlicher Bedeutung von den Eigenschaften der erfahrbaren Umwelt wird auch von S. L. Frank und Kollegen für das DSS-Modell hervorgehoben. Sie betrachten diese Abhängigkeit von den Strukturen der Welt als zentrale Quelle für die Herstellung von Systematizität (S. L. Frank et al., 2009, S. 376):

[...] the systematicity is developed robustly because it derives from the structure that is present in the world as well as the language used to describe that world. The origin of systematicity should be sought in the cognitive system's embeddedness in the world rather than in inherent properties of the system itself.

Aus dieser Perspektive erweitern wir unsere Auffassung von sprachlicher Bedeutung als kausalem Effekt auf mentale Zustände um die referenzielle Verankerung dieser Zustände: Die Bedeutung von Wörtern entspricht in dieser Konzeption den kausalen Effekten der Wörter auf die mentalen Zustände, deren Strukturen durch die analogen, multimodal verschränkten Erfahrungen des Systems mit der Welt geprägt sind.

Einerseits hilft die Konzeption von Bedeutung und Verstehen der Embodied Cognition dabei, konnektionistische Ansätze des Sprachverstehens zu konkretisieren und Modelleigenschaften zu deuten. Umgekehrt können die erwähnten konnektionistischen Konzep-

tionen von Repräsentation und Verarbeitung dabei helfen, vage Konzepte der Embodied Cognition zu konkretisieren und unbegründete Annahmen zu kritisieren (vgl. Kap. 9.2.2). Zunächst einmal erlaubt die Umsetzung in einem konnektionistischen Ansatz – wie etwa in unserem Modell SynSem – die Simulation von Prozessen wie etwa Selektion, Schematisierung, Kategorisierung und Prototypisierung. Die Fähigkeit, diese Prozesse zu leisten, wurde verschiedentlich als grundlegende Eigenschaft für ein vollfunktionales System kognitiver Verarbeitung im Ansatz der Embodied Cognition gefordert (z.B. Barsalou, 1999). Mit unserem Modell demonstrieren wir einen Vorschlag, wie diese Eigenschaften beschaffen sein könnten und wie sie sich auf die Verarbeitung auswirken. Übereinstimmend mit Joyce et al. (2003) können wir damit aber die Annahme symbolischer Mechanismen und Repräsentationsformate in Frage stellen, die Barsalou (1999) als Grundlage seines Ansatzes der perzeptuellen Symbole postuliert, und stellen dem eine subsymbolische Alternative als geeignetere Grundlage für die Theoriebildung über kognitive Repräsentationen im Ansatz der Embodied Cognition entgegen (vgl. auch Jackson & Sharkey, 1996; Reilly, 1995).

Des Weiteren wird für den Spracherwerb in Ansätzen der Embodied Cognition angenommen, dass verschiedene situational-semantische Aspekte durch verschiedene sprachliche Formen – Wörter wie grammatische Strukturen – indiziert werden (etwa Glenberg & Robertson, 2000; Glenberg & Kaschak, 2002; Bergen & Wheeler, 2009). Wie und mit welchen Eigenschaften diese Indizierung aber genau erfolgt, wird in den Ansätzen nicht spezifiziert. In unserem Modell simulieren wir die Indizierung als einfachen selbstorganisierten Assoziations-Prozess, in dem statistisch situationsübergreifend Korrespondenzen zwischen verschiedenen situationalen Aspekten und verschiedenen sprachlichen Formen abgeleitet werden.

Als letztes Beispiel für einen direkten Bezug nennen wir noch die Beschreibung des Verstehens im IEF-Ansatz (Zwaan, 2004) im Sinne multipler Constraint Satisfaction. Frühe Verstehensprozesse werden vage als Artikulationsprozess aktivierter Bedeutungsnetze beschrieben (vgl. Kap. 8.2.3). Artikulation entspricht dabei der gegenseitigen Beschränkung von Aktivierung durch verschiedene Informationen und der damit erfolgenden Konkretisierung der entwickelten Repräsentation. Artikulation kann sich vorwärts gerichtet auf neue Informationen auswirken, deren Bedeutung über den vorhandenen Kontext eingeschränkt wird, und rückwärts als Konkretisierung angenommener Bedeutungen durch neu hinzukommende Informationen. Genau diese angenommenen Auswirkungen können wir in unserem Modell beobachten, wenn etwa korrekt positionierte, aber unterspezifizierte Objekt-Repräsentationen durch die schließliche Nennung des Objekt-Namens konkretisiert werden. Es finden sich also viele Parallelen zwischen verschiedenen Ebenen des Modells und bislang nur theoretisch postulierten Mechanismen verschiedener Embodiment-Theorien der Sprachverarbeitung – womit das Modell eine reiche Quelle für Konkretisierungen der angenommenen Prozesse dieser Erklärungsansätze bietet.

### 10.4.4. Weiterentwicklungsmöglichkeiten für SynSem

SynSem ist kein abgeschlossenes Projekt. Wir haben in diesem Kapitel erst einen Anfang geschaffen, das prinzipielle Potenzial des Ansatzes ausgelotet und erste Hindernisse thematisiert. Von diesem Ausgangspunkt aus sind viele Erweiterungen denkbar, die interessante Modellierungsstudien eröffnen würden. Zudem gibt es viele Aspekte, in denen die

Performanz optimiert und die Erklärungskraft des Modells erweitert werden könnte. Im Folgenden benennen wir einige Möglichkeiten für zukünftige Projekte. Diese betreffen vor allem Modifikationen der visuellen und der sprachlichen Stimuli, die sich mitunter gegenseitig bedingen.

Eine naheliegende Möglichkeit zur Erweiterung des Modells, die das Feld möglicher Modellierungsgegenstände erheblich erweitern würde, betrifft die Gestaltung der visuellen Stimuli: Interessant wäre es zunächst, zwischen Objekt-Kategorien mit verschiedenen Vertretern zu unterscheiden, die unterscheidende Merkmale tragen – eine einfache Umsetzung könnte zum Beispiel über die Einführung großer und kleiner Objekte erfolgen (kleiner SLASH, zusammengesetzt aus zwei aktivierten Units, und großer SLASH, zusammengesetzt aus vier aktivierten Units). Dies würde erfordern, dass das Modell lernt, sowohl für Objekte als auch für Objekt-Merkmale Kategorien zu bilden und anhand kategorialer Merkmale Unterscheidungen zu treffen. Eine zweite Erweiterung der visuellen Stimuli betrifft die Eigenschaften der distributionalen Struktur. Wir konnten in Simulation SynSem I beim entkoppelten Verstehen feststellen, dass als Stellvertreter-Repräsentation der unvollständigen Objekt-Information leicht der SLASH aktiviert wurde. Wir haben vermutet, dass dies durch Ungleichmäßigkeiten der Verteilung im Trainings-Korpus begründet sein könnte. Solche Ungleichmäßigkeiten könnten eine weitere interessante Untersuchungsmöglichkeit eröffnen – etwa die kontrollierte Simulation von Einflüssen der Häufigkeitsunterschiede verschiedener Objekte auf die Verarbeitungsperformanz. Eine dritte Erweiterung der visuellen Stimuli betrifft die Möglichkeiten der Interaktion zwischen verschiedenen Objekten im Sinne kausaler Zusammenhänge. Dies würde aber eine nicht unwesentliche Veränderung der dargebotenen Situationen erfordern, die in irgendeiner Form eine situationale Veränderung enthalten müssten. Dies könnte beispielsweise über bewegliche Objekte realisiert werden.

Durch die oben vorgeschlagene Einführung von Objekt-Kategorien und unterscheidenden Merkmalen könnte eine Erweiterung der sprachlichen Stimuli eingeführt werden. Dies würde die Modellierung von Eigenschaften ermöglichen, die Gegenstand vieler psychologischer Studien zur Interaktion von Sprache und visueller Verarbeitung wurden – beispielsweise die Simulation der Verarbeitung restriktiver Relativsätze, die in der Studie von Tanenhaus et al. (1995) untersucht wurden (die Studie wurde in Kurzform in Kapitel 5.2.3 erläutert). Korrespondierend hierzu könnte bei entsprechend gestalteten Stimuli untersucht werden, ob für Sätze mit lokaler Ambiguität eine Disambiguierung der Struktur anhand des visuellen Kontexts erfolgt. Eine zweite Erweiterung der sprachlichen Stimuli könnte vergleichsweise unproblematisch durch eine differenziertere sprachliche Beschreibung der bereits vorhandenen Situationen erfolgen. Für diese könnte etwa unterschieden werden, ob die Objekte weit auseinander oder nahe beieinander liegen, oder ob sie sich berühren. Eine solche Erweiterung könnte zudem die Simulation der Verarbeitung von Präpositionen ermöglichen und damit eine Replizierung der Modellierungen von Regier (1996) erlauben. Mit der Erweiterung um Kontaktsituationen zwischen Objekten könnte das Lexikon unserer Miniatursprache zudem um weitere Verben ergänzt werden (z.B. *slash berührt backslash*). Eine dritte Erweiterung der sprachlichen Stimuli betrifft die Einführung von Pronomen. Dies würde eine interessante Untersuchungsmöglichkeit eröffnen, da die Auflösung der pronominalen Referenz im Modell SynSem anders als in unserem Ansatz AnaSRN (vgl. Kap. 6.6) innerhalb eines situationalen Kontexts erfolgen kann, und

deshalb weitere Aspekte der Auflösung berücksichtigt werden können. Eine Erweiterung, die vor allem aus der Perspektive des Embodied Language Processing sehr interessant sein könnte, wäre die Einführung der Negation: etwa *slash ist nicht links von backslash* (sondern *rechts* davon) oder *links von backslash ist nicht slash* (sondern *minus*). Verbunden mit der Paarung solcher Sätze mit kontrafaktischen visuellen Situationen, die zur Negation korrespondieren, könnte eine solche Modellierung eine alternative Erklärung der Repräsentation von Negation anbieten. Wir vermuten, dass die Negation eines Sachverhalts entgegen der Konzeption von Kaup et al. (2006) (vgl. Kap. 8.2.2) in Form expliziter Nicht-Aktivierung bei gleichzeitiger Aktivierung alternativer Sachverhalte erfasst wird. Zu beweisen wäre dabei, ob der negierte Sachverhalt bei dieser Form von Repräsentation dennoch präsent bleibt.

Speziell die Architektur betreffend haben wir an dieser Stelle keine konkreten Vorschläge für Modifikationen. Dennoch gäbe es gerade hier Weiterentwicklungsbedarf. Ein wichtiger Aspekt betrifft die Sprachproduktion: Wie wir in den jeweiligen Simulationsberichten zu diesem Aspekt diskutiert haben, ist die bisherige willkürliche und unstrukturierte Produktion von Wortsequenzen nur bedingt befriedigend. Allerdings haben wir bislang keinen Vorschlag anzubieten, wie eine Umsetzung der Verbesserung der Produktion aussehen könnte. Vor allem ist es schwierig, dabei auf die Manipulation der Stimuli zu verzichten – was etwa im Modell von Rohde (2002) durch explizite Hinweise zur syntaktischen Struktur geschieht. Nicht zuletzt wäre es ganz allgemein auf das Modell SynSem bezogen erstrebenswert, das oben ausführlich diskutierte Problem der Reduzierung und Prototypisierung ausschließlich implizit im Modell zu realisieren und damit jegliche Vorgabe über die Stimuli zu vermeiden. Ein erster Ansatz hierfür wurde mit SynSem II durch die Verlagerung der Prototypizität in die distributionale Struktur der Stimuli geleistet. Für eine sinnvolle Überführung der Reduzierung situationaler Aspekte hingegen fehlt bislang eine geeignete Umsetzung.

Wir glauben, dass das Projekt SynSem und der damit beschrittene Ansatz das Potenzial bietet, viele interessante Fragestellungen zu untersuchen, die Interaktion von Sprache und visueller Verarbeitung zu simulieren und theoretische Aspekte des Sprachverstehens aus der Perspektive der Embodied Cognition greifbarer und überhaupt erst modellierbar zu machen. Dementsprechend hoffen wir, dass die Ergebnisse dieser Arbeit nicht den Abschluss dieser Idee darstellen, sondern einen Ausgangspunkt für ihre Weiterentwicklung schaffen.



# 11. Schluss

In dieser Arbeit wurde in drei Teilen eine konnektionistische Perspektive auf verschiedene Aspekte des Sprachverstehens entwickelt. Abschließend wollen wir den Zusammenhang der besprochenen Themen skizzieren, und darstellen, wie die einzelnen Kapitel zur Beantwortung unserer Ausgangsfragen beitragen: Was bedeutet Sprachverstehen im Konnektionismus? Wie werden Referenzialität und Bedeutung in konnektionistischen Modellen erfasst? Und welche Eigenschaften, Faktoren und Mechanismen werden zur Erklärung menschlichen Sprachverstehens herangezogen? Im Anschluss fassen wir zentrale Ergebnisse der Arbeit zusammen. Dabei erläutern wir unsere Ergebnisse und Schlussfolgerungen aus den theoretischen Aufarbeitungen zu Konnektionismus und Embodied Cognition, aus dem empirischen Projekt AnaLoCo und aus den Modellierungsprojekten AnaSRN und SynSem – und diskutieren ihre Zusammenhänge und Konsequenzen.

## 11.1. Untersuchungsbereiche und Argumentationslinien

Übergreifendes Thema der Arbeit ist der Erwerb und die Verarbeitung von Referenzialität und Bedeutung bei der Satzverarbeitung. Wir haben hierfür zwei Kerngebiete identifiziert, in denen referenzielle Verarbeitung zur Herausforderung für die kognitive Verarbeitung wird: Auflösung von Koreferenzialität und Etablierung von Korrespondenz bei referenzieller Uneindeutigkeit. Anhand dieser Gegenstände haben wir eine konnektionistische Perspektive auf Sprachverstehensprozesse entwickelt. Hierbei haben wir zwei theoretische Themenfelder und drei praktisch umgesetzte Projekte behandelt: (a) Konnektionismus und Sprache, (b) Konnektionismus und Embodied Cognition, (c) die Modellierung von Anaphernauflösung, (d) die empirische Untersuchung von Anaphernauflösung und lokalen syntaktischen Kohärenzen, und (e) die Modellierung des situationalen Bedeutungserwerbs. Diese Themenfelder und Projekte haben wir in drei Hauptteile untergliedert.

In **Teil I** haben wir den Konnektionismus in der Tradition des Parallel Distributed Processing (McClelland & Rumelhart, 1986) und zentrale Eigenschaften hinsichtlich Repräsentation und Verarbeitung besprochen und eine Perspektive dieser kognitiven Theorie auf Debatten der menschlichen Sprachfähigkeit einschließlich Spracherwerb, -verarbeitung und -entwicklung entworfen. In der Folge haben wir ausführlich die zentrale Architektur der vorliegenden Arbeit diskutiert: Simple Recurrent Networks (Elman, 1990). Wir haben über deren Eigenschaften zentrale Faktoren identifiziert, die zur Erklärung der Sprachfähigkeit speziell betreffend Syntax und Grammatik herangezogen werden. Abschließend haben wir Erweiterungen der Architektur besprochen, die die Modellierung der referenziell-semantischen Verarbeitung und des Sprachverstehens ermöglichen (u.a. Rohde, 2002; Mayberry, 2003; Mayberry et al., 2009; S. L. Frank et al., 2009; Altmann & Mirkovic, 2009).

Auf den aufbereiteten konnektionistischen Grundlagen basierend haben wir in **Teil II** das Modellierungsprojekt *AnaSRN* vorgestellt, in dem wir mit erweiterten SRNs die Anaphernaufflösung bei der Satzverarbeitung simuliert haben. Vorbereitend wurden empirische Befunde und psycholinguistische Theorien zu den beiden Kernthemen lokale syntaktische Kohärenzen (Tabor et al., 2004; Konieczny & Müller, 2010) sowie Anaphernaufflösung (u.a. Badecker & Straub, 2002; Sturt, 2003; Runner et al., 2006; Kaiser et al., 2009) besprochen. Die Modellierungen legen verschiedene Eigenschaften für die syntaktische und anaphorische Verarbeitung nahe, die sich empirisch überprüfen lassen. Im Projekt *AnaLoCo* haben wir anhand empirischer Studien die Interaktion zwischen Anaphernaufflösung und LSK untersucht. Die Befunde legen den Einfluss lokaler Bindungsconstraints nahe, die mit globalen Constraints konkurrieren können.

In **Teil III** haben wir die Embodied Cognition als Rahmentheorie zur Erklärung sprachlicher Bedeutung vorgestellt (z.B. Barsalou, 1999; Glenberg & Kaschak, 2002; Zwaan, 2004) und ihre Kompatibilität mit konnektionistischen Theorien der Sprachverarbeitung untersucht (vgl. Plunkett et al., 1992; Joyce et al., 2003; Cangelosi, 2005; S. L. Frank et al., 2009). Während der Konnektionismus die Möglichkeit bietet, Theorien der Embodied Cognition zu modellieren und damit zu spezifizieren, bietet die Embodied Cognition einen Ansatz dafür, was in konnektionistischen Modellen als Bedeutung verstanden werden könnte. Im Projekt *SynSem* haben wir aus dieser Perspektive ein konnektionistisches Modell mit rekurrenter Dual-Route-Architektur zum Bedeutungserwerb und zur Satzverarbeitung vorgestellt, das das Problem der potenziellen Viele-zu-viele-Korrespondenz löst, und in dem sich Aufmerksamkeits- und Inferenzeffekte entwickeln.

## 11.2. Zentrale Ergebnisse und Schlussfolgerungen

Im Folgenden fassen wir zunächst kurz die aus den behandelten theoretischen Themenfeldern hervorgegangenen Annahmen zusammen. Anschließend skizzieren wir die zentralen Ergebnisse der umgesetzten Projekte, und benennen die daraus abgeleiteten Schlussfolgerungen.

**Konnektionismus und Sprachverstehen** Unser Ausgangspunkt für die Skizzierung einer Perspektive des Konnektionismus auf Sprachverstehensprozesse sind Modellierungen mit SRNs und SRN-basierten Modellen. Wir haben dabei zwischen SRNs mit Vorhersage-Aufgabe und mit der Aufgabe situationaler Assoziation unterschieden. SRNs zeichnen sich durch eine Reihe von Verarbeitungseigenschaften aus, die inhärent im Modell gegeben sind: Sie leisten Antizipation kommender Zustände aufgrund des gegebenen sequenziellen oder situationalen Kontexts. Sie leiten induktiv anhand statistischer Inferenzen aus der distributionalen Struktur des Inputs die Funktionen ab, die der Datenstruktur zugrunde liegen. Sie können Informationen unterschiedlicher Quellen integrieren und dabei die wechselseitigen Constraints dieser Quellen für die Ableitung versteckter struktureller Information nutzen. Allerdings wurden verschiedentlich Beschränkungen der Systematizität von SRNs festgestellt. Als Haupteinflussfaktoren auf die Performanz dieser Netzwerkarchitektur wurden distanzbasierte Interferenzen, Effekte der Frequenz und der Regularität von Input-Strukturen sowie Interferenzen lokaler Subsequenzen festgestellt und als Erklärungsansatz

auf Befunde der natürlichen Sprachverarbeitung übertragen. Auch für die Modellierung referenzieller Sprachverarbeitung sind SRNs geeignet: Sie können sequenzielle Strukturen erfassen, auf beliebige Zielstrukturen abbilden und beliebige Input-Strukturen integrieren. Aufgrund dieser Eigenschaften wurden sie zu einem Grundbestandteil komplexerer Architekturen für die Modellierung von Sprachverstehensprozessen. Für die Enkodierung der bedeutungshaften Zielstrukturen als Gegenstand des Verstehens kommen in den verschiedenen Modellen neben situationalen Repräsentationen vorwiegend propositionale Strukturen zum Einsatz. Diese werden verschiedentlich entweder als Input (Struktur der wahrgenommenen Welt), als Output (Struktur der mentalen Repräsentation) oder für die Enkodierung unabhängiger interner Repräsentationen von Bedeutung zugrunde gelegt. Die untersuchten Modelle teilen sich die Annahmen der verteilten Repräsentation und Verarbeitung konzeptuellen Wissens. Zugleich konnten wir aber feststellen, dass mit den Modellen sehr unterschiedliche Konzeptionen von Bedeutung und Verstehen verfolgt werden.

**Konnektionismus und Embodied Language Processing** Als mögliche theoretische Grundlage einer konnektionistischen Perspektive auf Bedeutung und Sprachverstehen haben wir den Forschungsrahmen der Embodied Cognition vorgestellt. In den Embodiment-Ansätzen zur Erklärung der allgemeinen Kognition wird Sprachverstehen weitgehend übereinstimmend als Simulation von analogen multimodalen Derivaten unserer Erfahrung mit der Welt beschrieben (z.B. Barsalou, 1999; Zwaan, 2004; Bergen & Wheeler, 2009). Im situierten Erwerb indizieren wir nicht-sprachliche, kategoriale, schematische Aspekte der Wahrnehmung von und Interaktion mit der Welt mit sprachlichen Formen. Beim Verstehen wird die Simulation durch Sprache geleitet, wir nutzen sprachliche Strukturen zur Konstruktion eines erfahrungsgeprägten mentalen Situationsmodells, das mit der beschriebenen Situation korrespondiert. Viele Grundannahmen, Konzepte und Termini der Embodied Cognition decken sich mit entsprechenden Grundlagen des Konnektionismus – etwa Analogizität, Prototypikalität, Kontextualität, Gradualität, Verarbeitungsbasiertheit, Constraint-Satisfaction und Emergenz als zentrale Eigenschaften von Erwerb, Repräsentation und Verarbeitung. Dementsprechend stellt der Konnektionismus eine geeignete Grundlage für die Modellierung theoretischer Konzepte der Embodied Cognition, und kann helfen, diese zu konkretisieren und auf Konsistenz zu prüfen. Allerdings muss dazu seine Entkörpertheit – das Fehlen materiell-physischer Merkmale und Eigenschaften – überwunden werden (vgl. Glenberg, 1997; Elman, 1999a). Umgekehrt kann die Embodied Cognition als theoretische Grundlage für eine konnektionistische Perspektive auf Bedeutung und Sprachverstehen dienen, die es ermöglicht, systematische Verarbeitungseigenschaften in einem genuin konnektionistischen Erklärungsrahmen zu implementieren, ohne dabei auf symbolverarbeitungsorientierte Hilfskonstrukte ausweichen zu müssen (S. L. Frank et al., 2009).

**Modellierung: Auflösung anaphorischer Referenz** Wir haben für die konnektionistische Modellierung referenzieller Sprachverarbeitung zwei Kerngebiete gewählt. Der erste Gegenstand ist die sprachinterne Referenzialität und die dabei zu lösende, vom sprachlich-strukturellen Kontext abhängige koreferenzielle Bindung. Wir haben im Projekt *AnaSRN* über eine einfache Erweiterung der SRN-Architektur ein Modell entwickelt, das uns erlaubt, die referenzielle und koreferenzielle Bindung bei der Satzverarbeitung

zu simulieren. Dabei können wir auch Eigenschaften von Sprachstrukturen untersuchen, die an der morphologisch-syntaktischen Oberfläche identisch, referenziell aber gänzlich unterschiedlich gebunden sind. Dies geht über das Spektrum hinaus, das mit normalen SRNs modelliert werden kann. Wir haben gezeigt, dass das Modell in der Lage ist, auch in komplexen syntaktischen Kontexten die korrekte koreferenzielle Bindung für Relativ-, Personal- und Reflexivpronomen zu erfassen. Die referenzielle Verarbeitung unseres Modells lässt sich dabei am ehesten als initialer Filter der in Betracht gezogenen Referenten deuten.

Für die Performanz der Wortvorhersage haben wir Effekte durch Frequenz, Regularität, Distanz und LSK gefunden, die bereits in früheren Studien gezeigt wurden (etwa Christiansen & Chater, 1999b; MacDonald & Christiansen, 2002; Konieczny, Ruh & Müller, 2009). Darüber hinaus konnten wir die Auswirkungen entsprechender Effekte für die Performanz der Referenzzuweisung zeigen (vgl. auch R. Frank et al., 2005). Dabei wurde die Ausprägung der Effekte mitunter dadurch bestimmt, welche grammatischen Eigenschaften die Trainings- und Test-Stimuli aufwiesen. Wir haben aus den Ergebnissen eine Reihe von Vorhersagen für empirische Studien abgeleitet: (a) Die Anaphernaufflösung erfolgt in einfachen Sätzen präziser oder schneller als in komplexen Sätzen, (b) Interferenz-Effekte werden vor allem durch alternative Referenten erwartet, die in grammatischen Kongruenz-Merkmalen mit dem korrekten Antezedens übereinstimmen. (c) LSK sollten sich auf die Anaphernaufflösung auswirken – das Agens der lokalen Subsequenz sollte verstärkt als Antezedens von Reflexivpronomen berücksichtigt und umgekehrt als Antezedens von Personalpronomen ausgeschlossen werden – und damit die Annahme der globalen Konsistenz der Interpretation bzw. der Bindungsregularitäten verletzen. (d) Zugleich sollten sich LSK-Effekte durch die unterschiedliche Abhängigkeit von strukturellen Bedingungen von Personal- und Reflexivpronomen unterschiedlich auf diese auswirken.

**Empirische Untersuchung: Anaphern und LSK** In der empirischen Studie *AnaLoCo* haben wir untersucht, ob die Einschätzungen aus der SRN-Modellierung für die Effekte auf die Anaphernaufflösung realistisch sind, und ob sich LSK-Effekte, die für andere Stimuli und Satzstrukturen gezeigt wurden (vgl. Tabor et al., 2004; Konieczny, 2005; Konieczny, Müller, Hachmann et al., 2009; Konieczny & Müller, 2010), auch für diesen Gegenstand finden. In einer Serie von drei Experimenten haben wir den Einfluss des strukturellen Phänomens LSK auf die Verarbeitung von Personal- und Reflexivpronomen untersucht. Im Self-Paced-Reading-Experiment konnten wir sowohl Effekte der LSK als auch des Anapherentyps zeigen, allerdings ergab sich keine Interaktion der beiden Faktoren. Zudem ließ sich mit dieser Methode keine Aussage darüber treffen, welchen Einfluss die LSK letztendlich auf die anaphorische Auflösung und die dabei berücksichtigten Referenten nimmt. Im ersten Visual-World-Experiment mit abgebildeten Referenten zeigten sich ebenfalls Effekte der LSK-Bedingung. Diese äußerten sich im Fall des Reflexivpronomens in höheren Fixationsanteilen auf dem vermeintlichen Agens der lokalen Subsequenz und im Fall des Personalpronomens in höheren Fixationsanteilen auf einem alternativen, nicht genannten Referenten. LSK scheinen also tatsächlich Einfluss auf die anaphorische Auflösung zu nehmen. Allerdings wiesen sowohl die visuellen als auch die sprachlichen Stimuli einige Eigenschaften auf, durch die eine alternative Interpretation der Daten nicht ausgeschlossen werden kann. Diese Einschränkungen haben wir in einem weiteren Visual-

World-Experiment mit abgebildeten Szenen kontrolliert. Sowohl für Reflexiv- als auch für Personalpronomen zeigten sich hier in der LSK-Bedingung Effekte einer erhöhten Fixationswahrscheinlichkeit für die Szenen, die die global inkonsistente Lesart der LSK abbilden. Während die Daten einige Fragen zur zeitlichen Auflösung der Verarbeitung und zu Unterschieden der verschiedenen pronominalen Formen offen ließen, findet sich der Effekt von LSK auf die Anaphernaufflösung bestätigt.

Mit diesen Ergebnissen der Studie können wir einige Erklärungsansätze zum Phänomen der LSK-Effekte in Frage stellen: So können die Effekte erstens nicht auf lexikalisch-statistische Effekte (vgl. u.a. R. L. Lewis et al., 2005; Gibson, 2006) zurückgeführt werden, da sich das entscheidende Wort (das Pronomen) in der Experimentalbedingung nicht von dem in der Kontrollbedingung unterscheidet, und die tiefergehende referenzielle Bindung nicht von der oberflächlichen lexikalischen Wahrscheinlichkeit abhängt. Zweitens können auch allgemeine Kontext-Effekte (vgl. van Dyke, 2007) ausgeschlossen werden, da die Referenten in den verschiedenen Bedingungen auf identischen Positionen stehen, und der Effekt von der syntaktischen Struktur abhängt. Drittens kann die Erklärung ausgeschlossen werden, die Effekte seien ein Artefakt der visuellen Induzierung, denn im Fall von Experiment 2 werden durch die visuellen Stimuli keine lokalen Lesarten nahegelegt. Damit stärken unsere Befunde die Annahme, dass LSK tatsächlich selbst interpretiert werden, und weiten den Wirkungsbereich dieses Phänomens auf die Anaphernverarbeitung aus.

**Modellierung: Situational-referenzieller Bedeutungserwerb** In der Modellierungsstudie *SynSem* haben wir einen Ansatz entwickelt, um situationalen Bedeutungserwerb und Prozesse der referenziellen Sprachverarbeitung in einer embodimentorientierten konnektionistischen Simulation umzusetzen. Das Modell kombiniert eine Dual-Route-Architektur (vgl. Plunkett et al., 1992; Cangelosi, 2005) mit einem SRN zu einem rekurrenten Dual-Route-SRN (ähnlich dem Modell von Dienes et al., 1999; Altmann & Mirkovic, 2009), das über einen gemeinsamen Hidden-Layer die Verarbeitung verschiedener Modalitäten bzw. Domänen vermittelt. Das Modell erlernt zwei Aufgaben: Wortvorhersage über eine sprachliche Route und Musterassoziation über eine visuelle Route. Als Trainings-Stimuli erhält es statisch dargebotene, einfache räumliche Objekt-Konstellationen auf einem retinalen Unit-Gitter sowie dazu korrespondierende beschreibende Sätze, die sequenziell Wort für Wort dargeboten werden. Beim Erwerb lernt das Modell über die Basis-Aufgaben hinaus, die nicht-sprachlichen Inhalte über die sprachlichen Formen zu indizieren, und Sprache somit nicht nur hinsichtlich ihrer rein sprachlichen, sondern auch hinsichtlich ihrer referenziellen Abhängigkeiten zu verarbeiten. Dabei werden erfolgreich Eins-zu-eins-Korrespondenzen zwischen einzelnen Wörtern sowie Satzzusammenhängen und situationalen Aspekten etabliert – das Modell löst selbstorganisiert das Problem der referenziellen Uneindeutigkeit. Drei Faktoren können als entscheidende Voraussetzungen für diese Leistung benannt werden: (a) Prototypisierungs- und Reduzierungsprozesse für die visuellen Stimuli, (b) situationsübergreifendes statistisches Lernen im Sinne einer situationalen Prototypisierung, und (c) die Informations-Integration sowie die Aufgaben-Koordination, die für die Erfüllung der Aufgaben über den gemeinsamen Hidden-Layer geleistet werden müssen. Die größte Schwachstelle des Modells besteht darin, dass die visuellen Reduzierungs- und Prototypisierungsprozesse entweder durch eine entsprechende Gestaltung der Target-Repräsentation oder in einer auto-assoziativen Variante durch entsprechende Vorgaben

für die distributionale Struktur der Input-Stimuli unterstützt werden mussten. In gewissen Grenzen können diese Vorgaben aber als realistische Vorannahmen für die Gestaltung der Stimuli aufgefasst werden.

Bei der Verarbeitung zeigt das Modell interessante Verhaltensmerkmale. Wird es ausschließlich mit sprachlichem Input getestet, so konstruiert es inkrementell ein Situationsmodell der zur Beschreibung korrespondierenden Situation. Dieser Verstehensprozess lässt sich im Sinne des Embodied Language Processing als Re-Enactment von Wahrnehmungsderivaten früherer Erfahrungen beschreiben. Erhält das Modell nur visuellen Input, so erzeugt es bei Erweiterung um eine Feedback-Schleife beschreibende Sätze. Allerdings wurde diese Aufgabe nur beschränkt erfolgreich absolviert. Bei der Verarbeitung mit vollständigem Input produziert das Modell ein Verhalten, das sich als situierte Sprachverarbeitung beschreiben lässt. Hierbei entwickeln sich als emergente Nebenprodukte der Verarbeitung, d.h. als nicht explizit trainierte Mechanismen, weitere Leistungsmerkmale und Eigenschaften des Modells: Es zieht aus den kombinierten Informationen frühzeitig Inferenzen und antizipiert die korrekte Interpretation des Satzes für die gegebene Situation. Dabei nutzt es den visuellen Input im Sinne modalitätenübergreifenden Primings für eine semantisch spezifizierte Wortvorhersage, und den sprachlichen Input im Sinne geleiteter visueller Aufmerksamkeit. Dieses Verhalten kann zu empirischen Befunden in Beziehung gesetzt werden, die eine starke Verschränkung sprachlicher und visueller Verarbeitung demonstrieren. Weitergehend eröffnet das Modell viel Raum für Interpretationen der Erwerbs- und Verarbeitungsmerkmale im Sinne des Embodied Language Processing. Zudem birgt es großes Potenzial für Weiterentwicklungen und die Untersuchung vielfältiger weiterer Fragestellungen. Zusammenfassend kann die Bedeutung von sprachlichen Elementen in unserer Konzeption als kausaler Effekt von Wörtern und grammatischen Strukturen auf die mentalen Zustände des kognitiven Systems beschrieben werden, dessen Strukturen durch die analogen, multimodalen Erfahrungen mit der Welt geprägt sind.

### 11.3. Perspektiven

Im Laufe der Promotion, in der diese Arbeit entstanden ist, haben wir sehr viele Projekte konzipiert, Experimente designt und Modellierungen durchgeführt. Wir haben unser Thema der referenziellen Sprachverarbeitung aus den dazu entwickelten Perspektiven nach bestem Wissen und Gewissen möglichst vielseitig und umfassend beleuchtet. Dennoch sollen die Projekte, die wir hier vorgestellt haben, nicht als abgeschlossene Untersuchungen betrachtet werden. Vielmehr hoffen wir, dass unsere Ansätze überzeugend genug waren, um als Anregung verstanden und weitergeführt zu werden. Wir hoffen auch, dass die vielen Anknüpfungspunkte ersichtlich wurden, die im Kleinen für detailliertere Folgefragestellungen und im Großen für Weiterentwicklungen der Modelle dienen können.

Für das Projekt *AnaSRN* sind viele Folgefragestellungen denkbar, die sich gezielt auf die Modellierung psycholinguistischer Daten und Befunde richten, die in den vergangenen zehn Jahren im Bereich der Erforschung anaphorischer Referenzen veröffentlicht wurden. Einfache Erweiterungen der Grammatik ermöglichen die Untersuchung der dort verwendeten syntaktischen Strukturen und Einflussfaktoren, und schaffen eine breitere Vergleichsbasis für die Performanz des Modells mit empirischen Daten. Im empirischen

Projekt *AnaLoCo* sind viele Fragen offengeblieben, und Details der gefundenen Effekte mussten eher spekulativ beantwortet werden. Kombinationen der bereits erstellten Materialien, gezielte Variationen von Details und verbesserte Kontrollstrukturen könnten helfen, das Bild der zeitlichen Auflösung der Verarbeitungsprozesse weiter zu klären und den Anteil verschiedener Einflussfaktoren an den Effekten präziser zu bestimmen. Das größte Potenzial für Weiterentwicklungen hat sicherlich das Projekt *SynSem*, für das wir im Projektkapitel eine ganze Reihe interessanter Erweiterungen der visuellen und sprachlichen Stimuli vorgeschlagen haben. Die vorgestellten Modelle verfügen bislang über eine sehr begrenzter Aussagekraft – sie zeigen lediglich die prinzipiellen Möglichkeiten des Ansatzes auf und demonstrieren grundlegende Leistungsmerkmale, die sich vage mit empirischen Befunden vergleichen lassen. Dabei versprechen die bisherigen Ergebnisse die Möglichkeit von Erweiterungen, die Vergleiche mit einem umfassenden Feld aktueller psycholinguistischer Forschung eröffnen.

Nicht zuletzt haben wir mit *SynSem* versucht, einen Ansatz für die Modellierung theoretischer Modelle und Annahmen der Embodied Cognition zu entwickeln. Für diesen Theorierahmen wurden bislang vor allem vage, grob skizzierte Modelle entworfen, um die Frage zu beantworten, welche kognitiven Prozesse für unsere Sprachverarbeitungsfähigkeit verantwortlich sind. Der Großteil der dazu vorgestellten empirischen Studien zeigt wenig mehr als generelle Hinweise darauf, dass beim Sprachverstehen multimodal-perzeptuelle und motorische Simulationen stattfinden. Die Entwicklung konkreter Modelle zum Embodied Language Processing könnte helfen, Lücken und Inkonsistenzen bestehender Theorien aufzudecken, vage formulierte Annahmen zu konkretisieren, und mit der Ableitung spezifischer Vorhersagen gezieltere empirische Studien anregen (vgl auch Barsalou, 2008). Dies ist unseres Erachtens der zwingende nächste Schritt für die Weiterentwicklung dieses Theoriefelds – und die konnektionistische Modellierung könnte hierfür ein geeignetes Instrument sein.



# Literaturverzeichnis

- Abney, S. (1989). Parsing by chunks. In C. Tenny (Hrsg.), *The MIT Parsing Volume*. MIT Press.
- Adamson, M. J. & Damper, R. I. (1999). B-RAAM: a connectionist model which develops holistic internal representations of symbolic structures. *Connection Science*, *11*, 41–71.
- Aisa, B., Mingus, B. & O'Reilly, R. (2008). The Emergent neural modeling system. *Neural Networks*, *21*, 1146–1152.
- Aizawa, K. (1997). Explaining systematicity. *Mind & Language*, *12*, 115–136.
- Allen, J. & Seidenberg, M. S. (1999). The emergence of grammaticality in connectionist networks. In B. MacWhinney (Hrsg.), *The emergence of language* (S. 115-152). Mahwah, NJ: Erlbaum.
- Altmann, G. T. M. (1998). Ambiguity in sentence processing. *Trends in Cognitive Science*, *2* (4), 146–152.
- Altmann, G. T. M. (1999). Thematic role assignment in context. *Journal of Memory and Language*, *41*, 124–145.
- Altmann, G. T. M. (2001). The language machine: Psycholinguistics in review. *British Journal of Psychology*, *92*, 129–170.
- Altmann, G. T. M. (2002). Learning and development in neural networks: the importance of prior experience. *Cognition*, *85*, 43–50.
- Altmann, G. T. M. (2006). History of Psycholinguistics. In K. Brown (Hrsg.), *The Encyclopedia of Language and Linguistics* (2. Aufl.). Elsevier.
- Altmann, G. T. M. & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, *73*, 247–264.
- Altmann, G. T. M. & Kamide, Y. (2007). The real-time mediation of visual attention by language and world knowledge: Linking anticipatory (and other) eye movements to linguistic processing. *Journal of Memory and Language*, *57* (4), 502–518.
- Altmann, G. T. M. & Kamide, Y. (2009). Discourse-mediation of the mapping between language and the visual world: Eye movements and mental representation. *Memory*, *111*, 55–71.
- Altmann, G. T. M. & Mirkovic, J. (2009). Incrementality and prediction in human sentence processing. *Cognitive Science*, *33*, 583–609.
- Altmann, G. T. M. & Steedman, M. (1988). Interaction with context during human sentence processing. *Cognition*, *30*, 198–238.
- Anderson, J. R. (2005). Human symbol manipulation within an integrated cognitive architecture. *Cognitive Science: A Multidisciplinary Journal*, *29* (3), 313–341.
- Anderson, J. R. & Lebiere, C. (1998). *Atomic components of thought*. Erlbaum.
- Anderson, M. (2003). Embodied cognition: a field guide. *Artificial Intelligence*, *149* (1), 91–130.

- Asudeh, A. & Keller, F. (2001). Experimental evidence for a predication-based binding theory. In M. Andronis, C. Ball, H. Elston & S. Neuvel (Hrsg.), *Papers from the 37th meeting of the chicago linguistic society* (Bd. 1: The Main Session, S. 1–14). Chicago.
- Auer, P. (1999). *Sprachliche Interaktion. Eine Einführung anhand von 22 Klassikern*. Tübingen: Max Niemeyer Verlag.
- Aziz-Zadeh, L., Wilson, S. M., Rizzolatti, G. & Iacoboni, M. (2006). Congruent embodied representations for visually presented actions and linguistic phrases describing actions. *Current Biology*, 16, 1–6.
- Baayen, R. H. (2008). *Analyzing linguistic data: a practical introduction to statistics using R*. Cambridge: Cambridge University Press.
- Baayen, R. H. (2010). languageR: Data sets and functions with „analyzing linguistic data: a practical introduction to statistics“ [Software-Handbuch]. Verfügbar unter <http://CRAN.R-project.org/package=languageR> (13.04.2010) (R package version 1.0)
- Baayen, R. H., Davidson, D. J. & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language* (59), 390–412.
- Bach, E., Brown, C. & Marslen-Wilson, W. (1986). Crossed and nested dependencies in german and dutch: a psycholinguistic study. *Language and Cognitive Processes*, 1, 249–262.
- Badecker, W. & Straub, K. (2002). The processing role of structural constraints on the interpretation of pronouns and anaphors. *Journal of Experimental Psychology: Learning, Memory & Cognition*, 28 (4), 748–769.
- Bailey, D. (1997). *A computational model of embodiment in the acquisition of action verbs*. Doctoral thesis, Computer Science Division, EECS Department, University of California, Berkeley, USA.
- Barr, D. J. (2008). Analyzing „visual world“ eyetracking data using multilevel logistic regression. *Journal of Memory and Language*, 59 (4), 457–474.
- Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, 22, 577–660.
- Barsalou, L. W. (2008). Grounded Cognition. *Annual Review of Psychology* (59), 617–45.
- Barsalou, L. W., Santos, A., Simmon, W. K. & Wilson, C. D. (2008). Language and simulation in conceptual processing. In M. D. Vega, A. M. Glenberg & A. Graesser (Hrsg.), *Symbols, embodiment, and meaning*. Oxford: Oxford University Press.
- Barsalou, L. W., Simmons, W. K., Barbey, A. K. & Wilson, C. D. (2003). Grounding conceptual knowledge in modality-specific systems. *Trends in Cognitive Sciences*, 7 (2), 84–91.
- Bates, D. M. (2005). Fitting linear mixed models in R. *R news*, 5, 27–30.
- Bates, E., Elman, J., Johnson, M., Karmiloff-Smith, A., Parisi, D. & Plunkett, K. (1999). Innateness and emergentism. In W. Bechtel & G. Graham (Hrsg.), *A companion to cognitive science* (S. 590–601). Oxford: Basil Blackwell.
- Baumann, P., Konieczny, L. & Hemforth, B. (2011). Coreference in Portuguese: pragmatic expectations based on alternative constructions and referring expressions. In B. Kozinov, A. Karmiloff-Smith & N. J. Nersessian (Hrsg.), *European Perspectives on Cognitive Science*. New Bulgarian University Press.

- Berg, G. (1992). A connectionist parser with recursive sentence structure and lexical disambiguation. In *Proceedings of the National Conference on Artificial Intelligence* (S. 32–32). John Wiley & Sons Ltd.
- Bergen, B. (2005). Mental simulation in spatial language processing. In B. G. Bara, L. Barsalou & M. Bucciarelli (Hrsg.), *Proceedings of the 27th Annual Conference of the Cognitive Science Society*. Mahwah, NJ: Erlbaum.
- Bergen, B. (2007). Experimental methods for simulation semantics. In M. Gonzalez-Marquez, I. Mittleberg, S. Coulson & M. Spivey (Hrsg.), *Methods in Cognitive Linguistics* (S. 277–301). Amsterdam: John Benjamins.
- Bergen, B. & Chang, N. (2003). Embodied construction grammar in simulation-based language understanding. In J.-O. Östman & M. Fried (Hrsg.), *Construction grammar(s): Cognitive grounding and theoretical extensions* (S. 147–190). Benjamin.
- Bergen, B., Chang, N. & Narayan, S. (2004). Simulated action in an embodied construction grammar. In *Proceedings of the 26th Annual Conference of the Cognitive Science Society*. Chicago, USA.
- Bergen, B., Lindsay, S., Matlock, T. & Narayanan, S. (2007). Spatial and linguistic aspects of visual imagery in sentence comprehension. *Cognitive Science: A Multidisciplinary Journal*, 31 (5), 733–764.
- Bergen, B. & Wheeler, K. (2005). Sentence understanding engages motor processes. In B. G. Bara, L. Barsalou & M. Bucciarelli (Hrsg.), *Proceedings of the 27th Annual Conference of the Cognitive Science Society*. Mahwah, NJ: Erlbaum.
- Bergen, B. & Wheeler, K. (2009). Grammatical aspect and mental simulation. *Brain and Language*, 112 (3), 150–158.
- Bever, T. G. (1970). The cognitive basis for linguistic structures. In J. R. Hayes (Hrsg.), *Cognition and the development of language*. New York: John Wiley & Sons.
- Bicknell, K. & Levy, R. (2009). A model of local coherence effects in human sentence processing as consequences of updates from bottom-up prior to posterior beliefs. In *Proceedings of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL HLT)*. Boulder, CO: Association for Computational Linguistics.
- Bodén, M. & Niklasson, L. F. (2000). Semantic systematicity and context in connectionist networks. *Connection Science*, 12 (2), 111–142.
- Boland, J., Tanenhaus, M. K., Carlson, G. N. & Garnsey, S. M. (1989). Lexical projection and the interaction of syntax and semantics in parsing. *Journal of Psycholinguistic Research*, 18 (6), 563–576.
- Borg, I. (2000). Explorative Multidimensionale Skalierung. *ZUMA How-to-Reihe* (1).
- Boroditsky, L. (2001). Does language shape thought? Mandarin and English speakers' conceptions of time. *Cognitive Psychology*, 43, 1–22.
- Boroditsky, L. (2003). Linguistic Relativity. In L. Nadel (Hrsg.), *Encyclopedia of Cognitive Science* (S. 917–921). Nature Publishing Group.
- Boroditsky, L., Ramscar, M. & Frank, M. C. (2001). The roles of body and mind in abstract thought. In J. D. Moore & K. Stenning (Hrsg.), *Proceedings of the 23rd Annual Meeting of the Cognitive Science Society*. Mahwah, NJ: Erlbaum.
- Borovsky, A. & Elman, J. (2006). Language input and semantic categories: a relation between cognition and early word learning. *Journal of Child Language*, 33, 759–790.

- Borreggine, K. L. & Kaschak, M. P. (2006). The Action-Sentence Compatibility Effect: It's all in the timing. *Cognitive Science: A Multidisciplinary Journal*, 30 (6), 1097–1112.
- Bortz, J. (1999). *Statistik für Sozialwissenschaftler* (3. Aufl.). Berlin: Springer.
- Boston, M. F., Hale, J. T., Kliegl, R., Patil, U. & Vasishth, S. (2008). Parsing costs as predictors of reading difficulty: an evaluation using the Potsdam Sentence Corpus. *Journal of Eye Movement Research*, 2, 1, 1–12.
- Botvinick, M. & Plaut, D. C. (2004). Doing without schema hierarchies: a recurrent connectionist approach to normal and impaired routine sequential action. *Psychological Review*, 111, 395–429.
- Bowers, J. S. (2009). On the biological plausibility of grandmother cells: Implications for neural network theories in psychology and neuroscience. *Psychological Review*, 116 (1), 220–251.
- Brakel, P. & Frank, S. L. (2009). Strong systematicity in sentence processing by simple recurrent networks. In N. Taatgen & H. vanRijn (Hrsg.), *Proceedings of the 31st Annual Meeting of the Cognitive Science Society*. Austin, TX.
- Büring, D. (2005). *Binding theory*. Cambridge, UK: Cambridge University Press.
- Brown-Schmidt, S., Byron, D. K. & Tanenhaus, M. K. (2005). Beyond salience: interpretation of personal and demonstrative pronouns. *Journal of Memory and Language*, 53, 292–313.
- Brunyé, T. T., Ditman, T., Mahoney, C. R., Augustyn, J. S. & Taylor, H. A. (2009). When you and I share perspectives. Pronouns modulate perspective taking during narrative comprehension. *Psychological Science*, 20 (1), 27–32.
- Bryant, J. (2004a). Recovering coherent interpretations using semantic integration of partial parses. In *Proceedings of the 3rd ROMAND workshop*.
- Bryant, J. (2004b). Scalable construction-based parsing and semantic analysis. In *Proceedings of the 2nd International Workshop on Scalable Natural Language Understanding (ScaNaLU 2004) at HLT-NAACL*.
- Brysbaert, M. (2007). „The language-as-fixed-effect fallacy“: some simple SPSS solutions to a complex problem (version 2.0). University of London: Royal Holloway.
- Bryson, J. (2005). *Embodiment vs. Memetics: Is building a human getting easier?* (Unpublished manuscript)
- Bullinaria, J. A. (1995). Modelling lexical decision: who needs a lexicon? In J. Keating (Hrsg.), *Neural computing research and applications iii (proceedings of the fifth irish neural networks conference)* (S. 62–69). Maynooth, Ireland: St. Patrick's College.
- Bullinaria, J. A. (1997). Analyzing the internal representations of trained neural networks. In A. Browne (Hrsg.), *Neural network analysis, architectures and algorithms*. Bristol: Institute of Physics Press.
- Burch, R. (2010). *Charles Sanders Peirce*. Stanford Encyclopedia of Philosophy. Verfügbar unter <http://plato.stanford.edu/entries/peirce/> (02.11.2010)
- Burgess, C. & Lund, K. (1997). Modelling parsing constraints with high-dimensional context space. *Language and Cognitive Processes*, 12 (2/3), 177–210.
- Bybee, J. L. & McClelland, J. L. (2005). Alternatives to the combinatorial paradigm of linguistic theory based on domain general principles of human cognition. *Review Literature And Arts Of The Americas*, 22 (2005), 381–410.
- Calvo Garzón, F. (2008). Connectionist semantics and the collateral information challenge. *Mind & Language*, 18, 77–94.

- Candliss, B. D. M. C. & Fiez, J. A. (2002). Success and failure in teaching the [r]-[l] contrast to Japanese adults: tests of a Hebbian model of plasticity and stabilization in spoken language perception. *Behavioral Neuroscience*, 2 (2), 89–108.
- Cangelosi, A. (2005). Approaches to grounding symbols in perceptual and sensorimotor categories. In H. Cohen & C. Lefebvre (Hrsg.), *Handbook of categorization in cognitive science* (S. 719–737). Elsevier Science.
- Cangelosi, A., Coventry, K. R., Rajapakse, R., Joyce, D., Bacon, A., Richards, L. et al. (2004). Grounding language in perception: a connectionist model of spatial terms and vague quantifiers. In *Ninth neural computation and psychology workshop*. Plymouth, UK.
- Cangelosi, A., Greco, A. & Harnad, S. (2000). From robotic toil to symbolic theft: Grounding transfer from entry-level to higher-level categories. *Connection Science*, 12, 143–162.
- Cangelosi, A., Greco, A. & Harnad, S. (2002). Symbol grounding and the symbolic theft hypothesis. In A. Cangelosi & D. Parisi (Hrsg.), *Simulating the evolution of language*. Springer-Verlag Publishing.
- Cangelosi, A. & Harnad, S. (2000). The adaptive advantage of symbolic theft over sensorimotor toil: Grounding language in perceptual categories. *Evolution of Communication*, 4 (1), 117–142.
- Cangelosi, A. & Parisi, D. (2002). Computer simulation: a new scientific approach to the study of language evolution. In A. Cangelosi & D. Parisi (Hrsg.), *Simulating the evolution of language*. Springer-Verlag Publishing.
- Cangelosi, A. & Riga, T. (2006). An embodied model for sensorimotor grounding and grounding transfer: experiments with epigenetic robots. *Cognitive Science*, 30, 673–689.
- Casasanto, D. (2006). Space for thinking. In V. Evans & P. Chilton (Hrsg.), *Language, Cognition and Space: the state of the art and new directions*. London: Equinox Publishing.
- Casasanto, D. (2007). When is a linguistic metaphor a conceptual metaphor? In V. Evans & S. Pourcel (Hrsg.), *New directions in cognitive linguistics* (S. 1–20).
- Casasanto, D. & Boroditsky, L. (2008). Time in the mind: using space to think about time. *Cognition*, 106 (2), 579–593.
- Chalmers, D. J. (1990). Syntactic transformations on distributed representations. *Connection Science*, 2 (1), 53–62.
- Chang, N., Feldman, J. & Narayanan, S. (2005). Structured connectionist models of language, cognition and action. *Progress in Neural Processing*, 16 (1996), 57–66.
- Charniak, E. (2001). A maximum-entropy-inspired parser. In *Proceedings of the first conference on North American chapter of the Association for Computational Linguistics* (S. 132–139).
- Chater, N. & Manning, C. D. (2006). Probabilistic models of language processing and acquisition. *Trends in Cognitive Sciences*, 10 (7), 335–44.
- Chomsky, N. (1957). *Syntactic structures*. The Hague: Mouton.
- Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge: MIT Press.
- Chomsky, N. (1975). *Reflections on language*. New York: Pantheon Books.
- Chomsky, N. (1980). *Rules and representations*. New York: Columbia University Press.
- Chomsky, N. (1981). *Lectures on government and binding*. Dordrecht: Foris.

- Chrisley, R. (2003). Embodied artificial intelligence. *Artificial Intelligence*, 149, 131–150.
- Chrisley, R. & Ziemke, T. (2003). *Embodiment*. Encyclopedia of Cognitive Science.
- Christiansen, M. H. (1994). *Infinite languages, finite minds. Connectionism, learning and linguistic structure*. Doctoral thesis, University of Edinburgh.
- Christiansen, M. H., Allen, J. & Seidenberg, M. S. (1998). Learning to segment speech using multiple cues: a connectionist model. *Language and Cognitive Processes*, 13, 221–268.
- Christiansen, M. H. & Chater, N. (1994). Generalization and connectionist language learning. *Mind and Language*, 9, 273–287.
- Christiansen, M. H. & Chater, N. (1999a). Connectionist natural language processing: the state of the art. *Cognitive Science*, 34 (4), 417–437.
- Christiansen, M. H. & Chater, N. (1999b). Toward a connectionist model of recursion in human linguistic performance. *Cognitive Science*, 23, 157–205.
- Christiansen, M. H. & Chater, N. (2001a). *Connectionist Psycholinguistics*. Westport, Connecticut: Ablex Publishing Corporation.
- Christiansen, M. H. & Chater, N. (2001b). Connectionist Psycholinguistics: Capturing the empirical data. *Trends in Cognitive Sciences*, 5 (2), 82–88.
- Christiansen, M. H. & Chater, N. (2008). Language as shaped by the brain. *Behavioral and Brain Sciences*, 31 (5), 489–509.
- Christiansen, M. H., Conway, C. M. & Curtin, S. (2005). Multiple-cue integration in language acquisition: a connectionist model of speech segmentation and rule-like behavior. In J. W. Minett & W. S.-Y. Wang (Hrsg.), *Language acquisition, change and emergence: Essay in evolutionary linguistics* (S. 205–249). Hong Kong: City University of Hong Kong Press.
- Christiansen, M. H., Dale, R. A. C., Ellefson, M. R., Conway, C. M., Cangelosi, A. & Parisi, D. (2002). The role of sequential learning in language evolution: Computational and experimental studies. In A. Cangelosi & D. Parisi (Hrsg.), *Simulating the evolution of language* (S. 165–187). Springer-Verlag Publishing.
- Christiansen, M. H. & Devlin, J. T. (1997). Recursive inconsistencies are hard to learn: a connectionist perspective on universal word order correlations. In *Proceedings of the 19th Annual Meeting of the Cognitive Science Society* (S. 113–118). Mahwah, NJ: Lawrence Erlbaum Associates.
- Christiansen, M. H. & Kirby, S. (2003). Language evolution: consensus and controversies. *Trends in Cognitive Science*, 7 (7), 300–307.
- Churchland, P. M. (1986). Some reductive strategies in cognitive neurobiology. *Mind*, 95, 279–309.
- Clark, H. H. (1973). The language-as-fixed-effect fallacy: a critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior*, 12, 335–359.
- Clifton, C. & Duffy, S. (2001). Sentence comprehension: Roles of linguistic structure. *Annual Review of Psychology*, 52, 167–196.
- Cohn, A. G., Gooday, J. M., Bennett, B. & Gotts, N. M. (1995). A logical approach to representing and reasoning about space. *Artificial Intelligence Review*, 9 (4–5), 255–259.
- Collins, A. M. & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, 82, 407–428.

- Collins, A. M. & Quillian, M. R. (1969). Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior*, 8 (2), 240–247.
- Cooper, R. M. (1974). The control of eye fixation by the meaning of spoken language: a new methodology for the real-time investigation of speech perception, memory, and language processing. *Cognitive Psychology* (6), 84–107.
- Copestake, A., Flickinger, D., Pollard, C. & Sag, I. A. (2005). Minimal recursion semantics: an introduction. *Research on Language & Computation*, 3 (4), 281–332.
- Coventry, K. R., Cangelosi, A., Rajapakse, R., Bacon, A., Newstead, S., Joyce, D. et al. (2005). Spatial prepositions and vague quantifiers: Implementing the functional geometric framework. In *Spatial cognition* (Bd. 4, S. 98–110). Springer.
- Coventry, K. R., Prat-Sala, M. & Richards, L. V. (2001). The interplay between geometry and function in the comprehension of ‘over’, ‘under’, ‘above’ and ‘below’. *Journal of Memory and Language*, 44, 376–398.
- Crain, S. & Steedman, M. (1985). On not being led up the garden-path: the use of context by the psychological parser. In D. Dowty, L. Karttunen & A. Zwicky (Hrsg.), *Natural language parsing*. Cambridge, UK: Cambridge University Press.
- Crocker, M. W., Pickering, M. & Clifton, C. (2000). *Architectures and mechanisms for language processing*. Cambridge, UK: Cambridge University Press.
- Dahan, D. & Tanenhaus, M. K. (2005). Looking at the rope when looking for the snake: Conceptually mediated eye movements during spoken-word recognition. *Psychonomic Bulletin & Review*, 12 (3), 453–459.
- Daneman, M. & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of verbal learning and verbal behavior*, 19 (4), 450–66.
- Davies, W. & Dubinsky, S. (2003). On extraction from nps. *Natural Language and Linguistic Theory*, 21, 1–37.
- Davis, M. H. (2002). Connectionist modelling of lexical segmentation and vocabulary acquisition. In P. Quinlan (Hrsg.), *Connectionist models of development: Developmental processes in real and artificial neural networks* (S. 151–187). Hove, United Kingdom: Psychology Press.
- de Saussure, F. (1967). *Grundfragen der allgemeinen Sprachwissenschaft* (3. Aufl.). Berlin: De Gruyter. (Übersetzung der französischen Originalausgabe von 1916)
- DeDe, G. (2010). Utilization of prosodic information in syntactic ambiguity resolution. *Journal of Psycholinguistic Research*, 39, 345–374.
- Dell, G., Chang, F. & Griffin, Z. (1999). Connectionist models of language production: Lexical access and grammatical encoding. *Cognitive Science: A Multidisciplinary Journal*, 23 (4), 517–542.
- DeLong, K. A., Urbach, T. P. & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience*, 8 (8), 1117–1121.
- Demberg, V. & Keller, F. (2009). A computational model of prediction in human parsing: Unifying locality and surprisal effects. In *Proceedings of the 29th Meeting of the Cognitive Science Society*.
- Desai, R. (2001). Bootstrapping in miniature language acquisition. In *Proceedings of the 4th International Conference on Cognitive Modeling* (S. 61–66). Hillsdale, NJ: Erlbaum.

- Desai, R. (2007). A model of frame and verb compliance in language acquisition. *Neurocomputing*, 70, 2273–2287.
- Devlin, J. T., Gonnerman, L. M., Andersen, E. S. & Seidenberg, M. S. (1998). Category-specific semantic deficits in focal and widespread brain damage: a computational account. *Journal of cognitive neuroscience*, 10 (1), 77–94.
- Dienes, Z., Altmann, G. T. M. & Gao, S.-J. (1999). Mapping across domains without feedback: a neural network model of transfer of implicit knowledge. *Cognitive Science*, 23, 53–82.
- Diessel, H. (2006). Demonstratives, joint attention, and the emergence of grammar. *Cognitive Linguistics*, 17, 463–489.
- Dominey, P. F. (2003). Learning grammatical constructions in a miniature language from narrated video events. In *Proceedings of the 25th Annual meeting of the Cognitive Science Society*. Boston.
- Dominey, P. F. & Boucher, J. D. (2005). Learning to talk about events from narrated video in a construction grammar framework. *Artificial Intelligence*, 167 (1-2), 31–61.
- Dominey, P. F. & Ramus, F. (2000). Neural network processing of natural language: I. sensitivity to serial, temporal, and abstract structure of language in the infant. *Language and Cognitive Processes*, 15 (1), 87–127.
- Dreyfus, H. L. (1979). *What computers can't do: the limits of artificial intelligence*. New York: Harper & Row.
- Dürscheid, C. (1991). *Modelle der Satzanalyse: Überblick und Vergleich* (Bde. Kölner Linguistische Arbeiten – Germanistik 16). Köln: Gabel.
- Eberhard, K. M., Spivey-Knowlton, M. J., Sedivy, J. C. & Tanenhaus, M. K. (1995). Eye movements as a window into real-time spoken language comprehension in natural contexts. *Journal of Psycholinguistic Research*, 24 (6), 409–436.
- Ellis, N. C. (1998). Emergentism, connectionism and language learning. *Language Learning*, 48 (4), 631–664.
- Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, 14, 179–211.
- Elman, J. L. (1991). Distributed representations, simple recurrent networks, and grammatical structure. *Machine Learning*, 7, 195–224.
- Elman, J. L. (1992). *Tlearn simulator*. Verfügbar unter <http://crl.ucsd.edu/innate/tlearn.html> (02.05.2006)
- Elman, J. L. (1993). Learning and development in neural networks: the importance of starting small. *Cognition*, 48, 71–99.
- Elman, J. L. (1995). Language as a dynamical system. In R. F. Port & T. van Gelder (Hrsg.), *Mind as motion* (S. 195–225). MIT press.
- Elman, J. L. (1998). Generalization, simple recurrent networks, and the emergence of structure. In M. A. Gernsbacher & S. Derry (Hrsg.), *Proceedings of the 20th Annual Conference of the Cognitive Science Society*. Mahway, NJ: Lawrence Erlbaum Associates.
- Elman, J. L. (1999a). Connectionism, artificial life, and dynamical systems: New approaches to old questions. In W. Bechtel & G. Graham (Hrsg.), *A companion to cognitive science* (S. 488–505). Oxford: Basil Blackwood.
- Elman, J. L. (1999b). Origins of language: a conspiracy theory. In B. MacWhinney (Hrsg.), *The emergence of language*. Hillsdale, NJ: Lawrence Erlbaum Associates.

- Elman, J. L. (2003). Generalization from sparse input. In *Proceedings of the 38th Annual Meeting of the Chicago Linguistic Society*.
- Elman, J. L. (2004). An alternative view of the mental lexicon. *Trends in Cognitive Science*, 8 (7), 301–306.
- Elman, J. L. (2005). Connectionist models of cognitive development: where next? *Trends in Cognitive Science*, 9 (3), 111–117.
- Elman, J. L. (2009). On the meaning of words and dinosaur bones: lexical knowledge without a lexicon. *Cognitive Science*, 33, 1–36.
- Elman, J. L., Bates, E., Johnson, M. H., Karmiloff-Smith, A., Parisi, D. & Plunkett, K. (1996). *Rethinking innateness: a connectionist perspective on development*. Cambridge, MA: MIT Press.
- Eurich, C. W. & Schwegler, H. (1997). Coarse coding: calculation of the resolution achieved by a population of large receptive field neurons. *Biological Cybernetics*, 76 (5), 357–363.
- Evans, V. & Green, M. (2006). *Cognitive linguistics: an introduction*. Edinburgh UP.
- Fanselow, G. & Felix, S. W. (1987). *Sprachtheorie: Eine Einführung in die Generative Grammatik* (Bd. 1 und 2). Francke.
- Farah, M. J. & McClelland, J. L. (1991). A computational model of semantic memory impairment: Modality specificity and emergent category specificity. *Journal of Experimental Psychology*, 120 (4), 339–357.
- Farkaš, I. & Crocker, M. W. (2006). Recurrent networks and natural language: exploiting self-organization. In *Proceedings of the 28th Annual Conference of the Cognitive Science Society*. Hillsdale, NJ: Lawrence Erlbaum.
- Farkaš, I. & Crocker, M. W. (2008). Syntactic systematicity in sentence processing with a recurrent self-organizing network. *Neurocomputing*, 71 (7-9), 1172–1179.
- Farkaš, I. & Li, P. (2002). Modeling the development of lexicon with a growing self-organizing map. In H. J. Caulfield et al. (Hrsg.), *Proceedings of the 6th Joint Conference on Information Sciences, Research Triangle Park, NC* (S. 553–556).
- Fedorenko, E., Gibson, E. & Rohde, D. (2006). The nature of working memory capacity in sentence comprehension: Evidence against domain-specific working memory resources. *Journal of Memory and Language*, 54 (4), 541–553.
- Feldman, J., Fianty, M. A., Goddard, N. H. & Lynne, K. J. (1988). Computing with structured connectionist networks. *Communications of the ACM*, 31 (2), 170–187.
- Feldman, J., Lakoff, G., Bailey, D., Narayanan, S., Regier, T. & Stolcke, A. (1996). L<sub>0</sub> – the first five years of an automated language acquisition project. *Artificial Intelligence Review*, 10 (1), 103–129.
- Feldman, J. & Narayanan, S. (2004). Embodied meaning in a neural theory of language. *Brain and Language*, 89 (2), 385–392.
- Ferreira, F., Bailey, K. G. D. & Ferraro, V. (2002). Good-enough representations in language comprehension. *Current Directions in Psychological Science*, 11 (1), 11–15.
- Ferreira, F. & Henderson, J. M. (1991). Recovery from misanalysis of garden-path sentences. *Journal of Memory and Language*, 30, 725–745.
- Ferreira, F. & Patson, N. D. (2007). The „good enough“ approach to language comprehension. *Language and Linguistics Compass*, 1 (1–2), 71–83.

- Fischer, I., Hennecke, F., Bannes, C. & Zell, A. (2002). *JavaNNS: Java Neural Network Simulator*. Universität Tübingen. Verfügbar unter <http://www.ra.cs.uni-tuebingen.de/software/JavaNNS/> (21.08.2006)
- Fischer, S. (2004). Optimal binding. *Natural Language & Linguistic Theory*, 22, 481–526.
- Fodor, J. A. (1976). *The language of thought*. Sussex: Harvester Press.
- Fodor, J. A. & Pylyshyn, Z. W. (1988). Connectionism and cognitive architecture: a critical analysis. In S. Pinker & J. Mehler (Hrsg.), *Connections and symbols* (S. 3–71). Cambridge, Mass.: MIT Press.
- Frank, R. & Mathis, D. (2007). Transformational networks. In *Proceedings of the Workshop on Psychocomputational Models of Human Language Acquisition* (S. 22–27).
- Frank, R., Mathis, D. & Badecker, W. (2005). *The acquisition of anaphora by simple recurrent networks*. Johns Hopkins University, Baltimore, Maryland. (Unpublished manuscript)
- Frank, S. L. (2006a). *A computational model of sentence comprehension: from words to situations*. (Unpublished manuscript)
- Frank, S. L. (2006b). Learn more by training less: systematicity in sentence processing by recurrent networks. *Connection Science*, 18 (3), 287–302.
- Frank, S. L. (2006c). *Modelling language comprehension without propositions*. (Unpublished manuscript)
- Frank, S. L. (2007). Book review: From molecule to metaphor: a neural theory of language (Feldman, 2006). *Computational Linguistics*, 33 (2), 259–261.
- Frank, S. L. (2009). Surprisal-based comparison between a symbolic and a connectionist model of sentence processing. In *Proceedings of the 31st Annual Conference of the Cognitive Science Society*.
- Frank, S. L. (2010). Uncertainty reduction as a measure of cognitive processing effort. In *Proceedings of the 2010 Workshop on Cognitive Modeling and Computational Linguistics*.
- Frank, S. L. & Haselager, W. F. G. (2006). Robust semantic systematicity and distributed representations in a connectionist model of sentence comprehension. In R. Sun & N. Miyake (Hrsg.), *Proceedings of the 28th Annual Conference of the Cognitive Science Society* (S. 226–231). Erlbaum.
- Frank, S. L., Haselager, W. F. G. & van Rooij, I. (2009). Connectionist semantic systematicity. *Cognition*, 110, 358–379.
- Frank, S. L., Koppen, M., Noordman, L. G. M. & Vonk, W. (2003a). *A model for knowledge-based pronoun resolution*.
- Frank, S. L., Koppen, M., Noordman, L. G. M. & Vonk, W. (2003b). Modeling knowledge-based inferences in story comprehension. *Cognitive Science: A Multidisciplinary Journal*, 27 (6), 875–910.
- Frank, S. L., Koppen, M., Noordman, L. G. M. & Vonk, W. (2007). Modeling multiple levels of text representation. In F. Schmalhofer & C. A. Perfetti (Hrsg.), *Higher level language processes in the brain: inference and comprehension processes* (S. 133–157). Mahwah, NJ: Erlbaum.
- Frank, S. L., Koppen, M., Noordman, L. G. M. & Vonk, W. (2008). World knowledge in computational models of discourse comprehension. *Discourse Processes*, 45, 429–463.

- Frazier, L. (1987). Sentence processing: a tutorial review. In M. Coltheart (Hrsg.), *Attention and performance* (S. 559–585).
- Frege, G. (1966). *Logische Untersuchungen*. Göttingen: Vandenhoeck und Ruprecht.
- Garagnani, M., Shtyrov, Y. & Pulvermüller, F. (2004). Effects of attention on what is known and what is not: MEG evidence for functionally discrete memory circuits. *Frontiers in Human Neuroscience*, 3 (10), 100–106.
- Garagnani, M., Wennekers, T. & Pulvermüller, F. (2008). A neuroanatomically-grounded hebbian learning model of attention-language interactions in the human brain. *European Journal of Neuroscience*, 27 (2), 492–513.
- Garson, J. (2010). *Connectionism*. Verfügbar unter <http://plato.stanford.edu/entries/connectionism/> (02.11.2010) (Stanford Encyclopedia of Philosophy)
- Gennari, S. P. & MacDonald, M. C. (2008). Semantic indeterminacy in object relative clauses. *Journal of Memory and Language*, 58, 161–187.
- Gennari, S. P. & MacDonald, M. C. (2009). Linking production and comprehension processes: the case of relative clauses. *Cognition*, 111, 1–23.
- Gernsbacher, M. A. (1990). *Language comprehension as structure building*. Erlbaum.
- Gibbs, R. W. J. (2006). Metaphor interpretation as embodied simulation. *Mind & Language*, 21 (3), 434–458.
- Gibson, E. (1998). Linguistic complexity: locality of syntactic dependencies. *Cognition*, 68 (1), 1–76.
- Gibson, E. (2000). The dependency locality theory: a distance-based theory of linguistic complexity. In Y. Miyashita, A. Marantz & W. O’Neil (Hrsg.), *Image, language, brain* (S. 95–126). Cambridge, Massachusetts: MIT press.
- Gibson, E. (2006). The interaction of top-down and bottom-up statistics in the resolution of syntactic category ambiguity. *Journal of Memory and Language*, 54, 363–388.
- Gibson, E. & Thomas, J. (1999). Memory limitations and structural forgetting: the perception of complex ungrammatical sentences as grammatical. *Language and Cognitive Processes*, 14 (3), 225–248.
- Gleitman, L. (1990). The structural sources of verb meanings. *Language Acquisition*, 1 (1), 3–55.
- Glenberg, A. M. (1997). What memory is for. *Behavioral and Brain Sciences*, 20, 1–55.
- Glenberg, A. M. & Kaschak, M. P. (2002). Grounding language in action. *Psychonomic Bulletin & Review*, 9, 558–565.
- Glenberg, A. M. & Robertson, D. A. (2000). Symbol grounding and meaning: a comparison of high-dimensional and embodied theories of meaning. *Journal of Memory and Language*, 43, 379–401.
- Goldberg, A. E. (2006). The inherent semantics of argument structure: the case of the English ditransitive construction. In D. Geeraerts (Hrsg.), *Cognitive Linguistics: Basic readings* (S. 401–437). Mouton de Gruyter.
- Golden, R. M. & Rumelhart, D. E. (1993). A parallel distributed processing model of story comprehension and recall. *Discourse Processes*, 16, 203–237.
- Gorrell, G. & Webb, B. (2006). Generalized hebbian algorithm for incremental latent semantic analysis. In *INTERSPEECH-2005* (S. 1325–1328).
- Graesser, A. C., Singer, M. & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review*, 101 (3), 371–395.

- Grosz, B., Joshi, A. K. & Weinstein, S. (1995). Centering: a framework for modeling the local coherence of discourse. *Computational Linguistics*, 21 (2), 203–225.
- Grumbach, A. (1996). Grounding symbols into perceptions. *Artificial Intelligence Review*, 10 (3), 131–146.
- Hachmann, W. (2008). *Exekutive Funktionen und Sprachkompetenz. Individuelle Unterschiede in der Verarbeitung komplexer Sätze*. Masters thesis, Universität Freiburg, Deutsches Seminar.
- Hachmann, W., Konieczny, L. & Müller, D. (2009). Individual differences in the processing of complex sentences. In *Proceedings of the 31st Annual Conference of the Cognitive Science Society*.
- Hacker, W., Handrick, S. & Veres, T. (1996). *Lesespannentest*. Universität Dresden. (Unpublished manuscript)
- Hadley, R. (1994). Systematicity in connectionist language learning. *Mind and Language*, 9 (3), 247–272.
- Hadley, R., Rotaru-Varga, A., Arnold, D. & Cardei, V. (2001). Syntactic systematicity arising from semantic predictions in a hebbian-competitive network. *Connection Science*, 13, 73–94.
- Hale, J. T. (2001). A probabilistic earley parser as a psycholinguistic model. *North American Chapter Of The Association For Computational Linguistics (NAACL)*, 1–8.
- Hale, J. T. (2003). The information conveyed by words in sentences. *Journal of Psycholinguistic Research*, 32 (2), 101–123.
- Hale, J. T. (2011). What a rational parser would do. *Cognitive Science*, 35 (3), 399–443.
- Halliday, M. A. K. & Hasan, R. (1976). *Cohesion in English*. London: Longman.
- Hammerton, J., Osborne, M., Armstrong, S. & Daelemans, W. (2002). Introduction to special issue on machine learning approaches to shallow parsing. *Journal of Machine Learning Research*, 2, 551–558.
- Hare, M. & Elman, J. L. (1995). Learning and morphological change. *Cognition*, 56, 61–98.
- Harnad, S. (1990). The symbol grounding problem. *Physica D*, 42, 335–346.
- Hemforth, B. & Konieczny, L. (2002). Sätze und Texte verstehen und produzieren. In J. Müsseler & W. Prinz (Hrsg.), *Lehrbuch Allgemeine Psychologie*. Spektrum Akademischer Verlag.
- Hemforth, B., Konieczny, L. & Scheepers, C. (2000). Syntactic attachment and anaphor resolution. the two sides of relative clause attachment. Cambridge University Press.
- Hemforth, B., Konieczny, L. & Strube, G. (1993). Incremental syntax processing and parsing strategies. In *Proceedings of the 15th Annual Conference of the Cognitive Science Society* (S. 539–545). Hillsdale, NJ: Erlbaum.
- Hemforth, B., Scheepers, C., Konieczny, L., Pynte, J. & Colonna, S. (2009). *Language specific preferences in anaphor resolution*. Rotterdam, NL.
- Hilario, M. (1997). An overview of strategies for neurosymbolic integration. In *Proceedings of the Workshop on Connectionist-Symbolic Integration: from Unified to Hybrid Approaches*. Montreal, CA.
- Hinton, G. E. (1986). Learning distributed representations of concepts. In *Proceedings of the 8th Annual Conference of the Cognitive Science Society* (S. 1–12). Hillsdale, NJ: Erlbaum.

- Hirschfeld, G. & Zwitserlood, P. (2010). How vision is shaped by language comprehension – top-down feedback based on low-spatial frequencies. *Brain Research*, 1377, 78–83.
- Hirschfeld, G., Zwitserlood, P. & Dobel, C. (2010). Effects of language comprehension on visual processing – MEG dissociates early perceptual and late N400 effects. *Brain and Language*, 116 (2), 91–96.
- Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6, 65–70.
- Holyoak, K. J. & Hummel, J. E. (2000). The proper treatment of symbols in a connectionist architecture. In E. Dietrich & A. B. Markman (Hrsg.), *Cognitive dynamics: Conceptual and representational change in humans and machines* (S. 229–263). Mahwah, NJ: Erlbaum.
- Hummel, J. E. & Holyoak, K. J. (2001). A process model of human transitive inference. In M. Gattis (Hrsg.), *Spatial schemas in abstract thought* (S. 279–305). Cambridge, MA: MIT press.
- Hummel, J. E., Holyoak, K. J., Green, C., Doumas, L. A. A., Devnich, D., Kittur, A. et al. (2004). A solution to the binding problem for compositional connectionism. In S. D. Levy & R. Gayler (Hrsg.), *Compositional Connectionism in Cognitive Science: Papers from the AAAI Fall Symposium* (S. 31–34). Menlo Park, CA: AAAI Press.
- Jackson, S. A. & Sharkey, N. E. (1996). Grounding computational engines. *Artificial Intelligence Review*, 10, 65–82.
- Jacobs, R. A., Jordan, M. I. & Barto, A. G. (1990). Task decomposition through competition in a modular connectionist architecture: the what and where vision tasks. *Cognitive Science*, 15, 219–250.
- Jaeger, H. (2003). Adaptive nonlinear system identification with echo state networks. In S. Becker, S. Thrun & K. Obermayer (Hrsg.), *Advances in neural information processing systems* (Bd. 15, S. 593–600). Cambridge, MA: MIT Press.
- Jaeger, T. F. (2008). Categorical data analysis: away from anovas (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, 59 (4), 434–446.
- Jaeger, T. F., Fedorenko, E. & Gibson, E. (submitted). *Anti-locality in English: Consequences for theories of sentence comprehension*. (Unpublished manuscript)
- James, D. L. & Miikkulainen, R. (1995). SARDNET: a self-organizing feature map for sequences. In G. Tesauro, D. S. Touretzky & T. K. Leen (Hrsg.), *Advances in neural information processing systems* (Bd. 7, S. 577–584).
- Joanisse, M. & Seidenberg, M. (2003). Phonology and syntax in specific language impairment: Evidence from a connectionist model. *Brain and Language*, 86, 40–56.
- Johnson, K. (2004). On the systematicity of language and thought. *Journal of Philosophy*, 101 (3), 111–139.
- Johnson-Laird, P. N. (1980). Mental models in cognitive science. *Cognitive Science*, 4, 71–115.
- Johnson-Laird, P. N. (2005). Mental Models and thought. In K. J. Holyoak & R. G. Morrison (Hrsg.), *The Cambridge Handbook of Thinking and Reasoning* (S. 185–208). Cambridge: Cambridge University Press.
- Jordan, M. I. (1986). *Serial order: a parallel distributed processing approach* (Bericht Nr. 8604). San Diego, CA: Institute for Cognitive Science, University of California.

- Joyce, D. W., Richards, L. V., Cangelosi, A. & Coverntry, K. R. (2003). On the foundations of Perceptual Symbol Systems: Specifying embodied representations via connectionism. In F. Detje, D. Dörner & H. Schaub (Hrsg.), *The Logic of Cognitive Systems. Proceedings of the 5th International Conference on Cognitive Modeling* (S. 147–152). Universitätsverlag Bamberg.
- Jurafsky, D. (2003). Probabilistic modeling in psycholinguistics: Linguistic comprehension and production. In R. Bod, J. Hay & S. Jannedy (Hrsg.), *Probabilistic Linguistics*. MIT press.
- Just, M. A. & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, 99 (1), 122–149.
- Just, M. A., Carpenter, P. A. & Woolley, J. D. (1982). Paradigms and processes in reading comprehension. *Journal of Experimental Psychology: General*, 111, 228–238.
- Kaiser, E., Runner, J. T., Sussman, R. S. & Tanenhaus, M. K. (2009). Structural and semantic constraints on the resolution of pronouns and reflexives. *Cognition*, 112, 55–80.
- Kamide, Y., Scheepers, C. & Altmann, G. T. M. (2003). Integration of syntactic and semantic information in predictive processing: Cross-linguistic evidence from German and English. *Journal of Psycholinguistic Research*, 32, 37–55.
- Kaplan, F., Oudeyer, P.-Y. & Bergen, B. (2007). Computational models in the debate over language learnability. *Infant and Child Development*, 17, 55–80.
- Kaplan, R. & Bresnan, J. (1982). Lexical-functional grammar: a formal system for grammatical representation. In J. Bresnan (Hrsg.), *The mental representation of grammatical relations* (S. 173–281). Cambridge, MA: MIT Press.
- Kasabov, N., Postma, E. & van den Herik, J. (2000). Avis: a connectionist based framework for integrated auditory and visual information processing. *International Journal on Information Sciences*, 123, 127–148.
- Kaschak, M. P. & Glenberg, A. M. (2000). Constructing meaning: the role of affordances and grammatical constructions in sentence comprehension. *Journal of Memory and Language*, 43, 508–529.
- Kaschak, M. P., Madden, C. J., Therriault, D. J., Yaxley, H. R., Aveyard, M., Blanchard, A. A. et al. (2005). Perception of motion affects language processing. *Cognition*, 94, 79–89.
- Kaup, B. (2006). What psycholinguistic negation research tells us about the nature of the working-memory representations utilized in language comprehension. In H. Pishwa (Hrsg.), *Language and memory: aspects of knowledge representation* (S. 313–355). Berlin: Mouton de Gruyter.
- Kaup, B., Lüdtke, J. & Zwaan, R. A. (2005). Effects of negation, truth value, and delay on picture recognition after reading affirmative and negative sentences. In B. G. Bara, L. Barsalou & M. Bucciarelli (Hrsg.), *Proceedings of the 27th Annual Conference of the Cognitive Science Society* (S. 1114–1119). Mahwah, NJ: Erlbaum.
- Kaup, B., Lüdtke, J. & Zwaan, R. A. (2006). Processing negated sentences with contradictory predicates: is a door that is not open mentally closed? *Journal of Pragmatics*, 38, 1033–1050.
- Kaup, B., Yaxley, R. H., Lüdtke, C. J. & Zwaan, R. A. (2007). Experiential simulations of negated text information. *Quarterly Journal of Experimental Psychology*, 60, 976–990.

- Kaup, B. & Zwaan, R. A. (2003). Effects of negation and situational presence on the accessibility of text information. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 29 (3), 439 – 446.
- Keller, F. & Asudeh, A. (2001). Constraints on linguistic coreference: structural vs. pragmatic factors. In J. D. Moore & K. Stenning (Hrsg.), *Proceedings of the 23rd Annual Conference of the Cognitive Science Society* (S. 483–488). Mahawah, NJ: Lawrence Erlbaum.
- Kello, C. T. & Plaut, D. C. (2003). Strategic control over rate of processing in word reading: a computational investigation. *Journal of Memory and Language*, 48, 207–232.
- Kennison, S. M. (2003). Comprehending the pronouns her, him, and his: Implications for theories of referential processing. *Journal of Memory and Language*, 49, 335–352.
- King, J. & Just, M. A. (1991). Individual differences in syntactic processing: the role of working memory. *Language*, 30, 580–602.
- Kintsch, W. (1998). *Comprehension: a paradigm for cognition*. New York, US: Cambridge University Press.
- Kintsch, W. & van Dijk, T. (1978). Toward a model of text comprehension and production. *Psychological Review*, 85, 363–394.
- Kleiber, G. (1993). *Prototypensemantik: Eine Einführung*. Tübingen: Narr.
- Knoeferle, P. & Crocker, M. W. (2006). The coordinated interplay of scene, utterance, and world knowledge: evidence from eye-tracking. *Cognitive Science*, 30 (3), 481–529.
- Kohonen, T. (1990). The self-organizing map. In *Proceedings of the IEEE* (Bd. 78, S. 1464–1480).
- Konieczny, L. (1996). *Human sentence processing: a semantics-oriented parsing approach*. Freiburg: IIG-Berich.
- Konieczny, L. (2000). Locality and parsing complexity. *Journal of Psycholinguistic Research*, 29 (6), 627–645.
- Konieczny, L. (2005). The psychological reality of local coherences in sentence processing. In *Proceedings of the 27th Annual Conference of the Cognitive Science Society* (S. 1178–1183).
- Konieczny, L. & Döring, P. (2003). Anticipation of clause-final heads. evidence from eye-tracking and SRNs. In P. P. Slezak (Hrsg.), *Proceedings of the 4th International Conference on Cognitive Science* (S. 330–335). University of New South Wales, Sydney, Australia.
- Konieczny, L., Haser, V., Müller, D., Weldle, H., Wolfer, S. A., Baumann, P. et al. (2010). Grounding of anaphora in pointing gestures. In *Proceedings of the 23d Annual Meeting of the CUNY Conference on Human Sentence Processing*. New York.
- Konieczny, L., Hemforth, B., Scheepers, C. & Strube, G. (1997). The role of lexical heads in parsing: Evidence from german. *Language and Cognitive Processes*, 12, 307–348.
- Konieczny, L. & Müller, D. (2007). Local syntactic coherence interpretation in spoken language comprehension. In *Proceedings of the 20th Annual CUNY Conference on Human Sentence Processing*. San Diego, CA.
- Konieczny, L. & Müller, D. (2010). Das sprachliche Arbeitsgedächtnis: Kapazität oder Erfahrung. *Psychologische Rundschau*, 61 (1), 43–50.
- Konieczny, L., Müller, D., Baumann, P., Hachmann, W. & Wolfer, S. (2009). Local syntactic coherence interpretation, and how prosody modulates it. In *Proceedings of*

- the 22nd Annual CUNY Conference on Human Sentence Processing*. Davis, CA.
- Konieczny, L., Müller, D., Hachmann, W., Schwarzkopf, S. & Wolfer, S. (2009). Local syntactic coherence interpretation. evidence from a visual world study. In N. Taatgen & H. vanRijn (Hrsg.), *Proceedings of the 31st Annual Meeting of the Cognitive Science Society* (S. 1133–1138). Austin, TX.
- Konieczny, L. & Ruh, N. (2003). *What's in an error? A reply to MacDonald & Christiansen (2002)*. (Unpublished manuscript)
- Konieczny, L., Ruh, N. & Müller, D. (2009). What's in an error? A closer look at SRNs processing relative clauses. In H. D. Peebles & R. P. Cooper (Hrsg.), *Proceedings of the 9th International Conference on Cognitive Modeling*. Manchester, UK.
- Konieczny, L. & Weldle, H. (2009). Embodied perspectives. In P. Auer & M. Hilpert (Hrsg.), *Workshop of the Freiburg Institute for Advanced Studies (FRIAS): Grammar, Space and Cognition*. Freiburg.
- Konieczny, L., Weldle, H., Haser, V., Müller, D., Wolfer, S. A. & Baumann, P. (2010). Anaphern sind in Zeigegeesten verankert - der Pointing-Compatibility Effect. In C. Frings, A. Mecklinger, D. Wentura & H. Zimmer (Hrsg.), *Beiträge zur 52. Tagung experimentell arbeitender Psychologen*. Saarbrücken.
- Konieczny, L., Weldle, H., Wolfer, S., Müller, D. & Baumann, P. (2010). Anaphora and local coherences. In *Proceedings of the 32nd Annual Conference of the Cognitive Science Society*. Portland, Oregon.
- Konieczny, L., Weldle, H., Wolfer, S., Müller, D., Baumann, P., Feldmeth, A. K. et al. (2009). Local coherences interfere with online anaphora resolution. In *Proceedings of the 15th Annual Conference on Architectures and Mechanisms for Language Processing (AMLaP)*. Barcelona, Spain.
- Laakso, A. & Cottrell, G. (2000). Content and cluster analysis: assessing representational similarity in neural systems. *Philosophical Psychology*, 13 (1), 47–76.
- Lakoff, G. (1987). *Women, fire, and dangerous things: what categories reveal about the mind*. Chicago: University of Chicago Press.
- Lakoff, G. & Johnson, M. (1980). *Metaphors we live by*. Chicago: University of Chicago Press.
- Lakoff, G. & Johnson, M. (1999). *Philosophy in the flesh: the embodied mind and its challenge to western thought*. New York: Basic Books.
- Landauer, T. K., Foltz, P. W. & Laham, D. (1998). Introduction to Latent Semantic Analysis. *Discourse Process*, 25, 259–284.
- Langacker, R. W. (1987). *Foundations of cognitive grammar (vol. 1): theoretical prerequisites*. Stanford, CA: Stanford University Press.
- Lehmann, C. (2007). *Semantik. Theoretische und empirische Bereiche*. Universität Erfurt. Verfügbar unter [http://www.christianlehmann.eu/ling/lg\\_system/sem/index.html](http://www.christianlehmann.eu/ling/lg_system/sem/index.html) (13.07.2010)
- Levinson, S. C. (1996). Relativity in spatial conception and description. In J. J. Gumperz & S. C. Levinson (Hrsg.), *Rethinking linguistic relativity* (S. 177–202). Cambridge, UK: Cambridge University Press.
- Levinson, S. C. (2003). *Space in language and cognition: explorations in cognitive diversity*. Cambridge, UK: Cambridge University Press.
- Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, 106, 1126–1177.

- Levy, R. & Andrew, G. (2006). Tregex and Tsurgeon: tools for querying and manipulating tree data structures. In *Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC)*.
- Levy, R., Fedorenko, E., Breen, M. & Gibson, T. (submitted). The processing of extraposed structures in English. *Journal of Memory and Language*.
- Levy, S., Melnik, O. & Pollack, J. (2000). Infinite RAAM: a principled connectionist basis for grammatical competence. In *Proceedings of the 22nd Annual Meeting of the Cognitive Science Society* (S. 298–303). Mahwah, NJ: Lawrence Erlbaum Associates.
- Lewis, J. & Elman, J. L. (2001a). A connectionist investigation of linguistic arguments from the poverty of the stimulus. learning the unlearnable. In J. D. Moore & K. Stenning (Hrsg.), *Proceedings of the 23th Annual Conference of the Cognitive Science Society* (S. 552–557). Mahwah, NJ: Lawrence Erlbaum.
- Lewis, J. & Elman, J. L. (2001b). Learnability and the statistical structure of language: Poverty of stimulus arguments revisited. In B. Skarabela, S. Fish & A. H. J. Do (Hrsg.), *Proceedings of the 26th Annual Boston University Conference on Language Development* (S. 359–370). Somerville, MA: Cascadilla Press.
- Lewis, R. L. & Vasishth, S. (2006). Computational principles of working memory in sentence comprehension. *Trends in Cognitive Science*, 10 (10), 447–454.
- Lewis, R. L., Vasishth, S. & van Dyke, J. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science*, 29, 1–45.
- Li, P., Farkaš, I. & MacWhinney, B. (2004). Early lexical development in a self-organizing neural network. *Neural Networks*, 17, 1345–1362.
- Lightfoot, D. (1999). *The development of language: acquisition, change, and evolution*. Malden, MA: Blackwell Publishers.
- Louwerse, M. M. (2007). Symbolic or embodied representations: a case for symbol interdependency. In T. Landauer, D. McNamara, S. Dennis & W. Kintsch (Hrsg.), *Handbook of Latent Semantic Analysis* (S. 107–120). Mahwah, NJ: Erlbaum.
- Maass, W., Natschläger, T. & Markram, H. (2002). Real-time computing without stable states: a new framework for neural computation based on perturbations. *Neural Computation*, 14 (11), 2531–2560.
- MacDonald, M. C. (1999). Distributional information in language comprehension, production, and acquisition: three puzzles and a moral. In B. MacWhinney (Hrsg.), *The emergence of language* (S. 177–196).
- MacDonald, M. C. & Christiansen, M. H. (2002). Reassessing working memory: comment on Just and Carpenter (1992) and Waters and Caplan (1996). *Psychological Review*, 109 (1), 35–54.
- MacDonald, M. C., Pearlmutter, N. J. & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review*, 101 (4), 676–703.
- MacDonald, M. C. & Seidenberg, M. S. (2006). Constraint satisfaction accounts of lexical and sentence comprehension. In M. J. Traxler & M. A. Gernsbacher (Hrsg.), *Handbook of Psycholinguistics* (2. Aufl., S. 581–611). London, UK: Elsevier.
- MacWhinney, B. (1998). Models of the emergence of language. *Annual Review of Psychology*, 49, 199–227.
- MacWhinney, B. (1999a). *The emergence of language*. Mahwah, NJ: Lawrence Erlbaum Associates.

- MacWhinney, B. (1999b). The emergence of language from embodiment. In B. MacWhinney (Hrsg.), *The emergence of language*. Mahwah, NJ: Lawrence Erlbaum Associates.
- MacWhinney, B. (2000). *The CHILDES project: tools for analyzing talk* (3. Aufl.). Mahwah, NJ: Lawrence Erlbaum Associates.
- MacWhinney, B. (2001). Language emergence. In P. Burmeister, P. Thorsten & A. Rohde (Hrsg.), *An integrated view of language development* (S. 17–42). Trier: Wissenschaftlicher Verlag.
- MacWhinney, B. (2008). How mental models encode embodied linguistic perspectives. In L. Klatzky, B. MacWhinney & M. Behrmann (Hrsg.), *Embodiment, ego-space, and action* (S. 365–406). New York: Psychology Press.
- Maechler, M., Rousseeuw, P., Struyf, A. & Hubert, M. (2005). *Cluster Analysis basics and extensions*. (Unpublished manuscript)
- Mangasser-Wahl, M. (2000). *Prototypentheorie in der Linguistik. Anwendungsbeispiele – Methodenreflexion – Perspektiven*. Tübingen: Stauffenburg.
- Marconi, D. (1996a). On the referential competence of some machines. *Artificial Intelligence Review*, 10, 21–35.
- Marconi, D. (1996b). Work on the integration of language and vision at the University of torino. *Artificial Intelligence Review*, 10, 15–20.
- Marcus, G. F. (1998). Rethinking eliminative connectionism. *Cognitive Psychology*, 37, 243–282.
- Marcus, M. P., Santorini, B. & Marcinkiewicz, M. A. (1993). Building a large annotated corpus of English: the Penn Treebank. *Computational Linguistics*, 19 (2), 313–330.
- Markman, A. B. & Brendl, C. M. (2005). Constraining theories of Embodied Cognition. *Psychological Science*, 16 (1), 6–10.
- Markman, A. B. & Dietrich, E. (2000). Extending the classical view of representation. *Trends in Cognitive Science*, 4, 470–475.
- Marr, D. (1982). *Vision: a computational investigation into the human representation and processing of visual information*. San Francisco: W. H. Freeman.
- Marrone, P. (2007). *Java Object Oriented Neural Engine (JOONE)*. Verfügbar unter <http://www.jooneworld.com> (13.06.2008)
- Mayberry, M. R. (2003). *Incremental nonmonotonic parsing through semantic self-organization*. Doctoral thesis, University of Texas at Austin.
- Mayberry, M. R., Crocker, M. W. & Knoeferle, P. (2005). A connectionist model of sentence comprehension in visual worlds. In *Proceedings of the 27th Annual Conference of the Cognitive Science Society*. Vancouver, CA.
- Mayberry, M. R., Crocker, M. W. & Knoeferle, P. (2009). Learning to attend: a connectionist model of situated language comprehension. *Cognitive Science*, 33 (3), 449–496.
- Mayberry, M. R. & Miikkulainen, R. (1999). SardSRN: a neural network shift-reduce parser. In *Proceedings of the 16th Annual Joint Conference on Artificial Intelligence*. Stockholm, Sweden.
- Mayberry, M. R. & Miikkulainen, R. (2003). Incremental nonmonotonic parsing through semantic self-organization. In *Proceedings of the 25th Annual Conference of the Cognitive Science Society*. Boston, MA.
- McClelland, J. L. (1989). Parallel distributed processing: implications for cognition and development. In R. G. M. Morris (Hrsg.), *Parallel distributed processing:*

- Implications for psychology and neurobiology.* (S. 8–45). Oxford: Clarendon Press.
- McClelland, J. L., Botvinick, M. M., Noelle, D. C., Plaut, D. C., Rogers, T. T., Seidenberg, M. S. et al. (2010). Letting structure emerge: connectionist and dynamical systems approaches to understanding cognition. *Trends in Cognitive Sciences*, 14, 348–356.
- McClelland, J. L. & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, 18 (1), 1–86.
- McClelland, J. L. & Kawamoto, A. H. (1986). Mechanisms of sentence processing: assigning roles to constituents of sentences. In J. L. McClelland & D. E. Rumelhart (Hrsg.), *Parallel distributed processing: explorations in the microstructure of cognition. volume 2: psychological and biological models* (S. 318–362). Cambridge, MA: MIT Press.
- McClelland, J. L. & Patterson, K. (2002). Rules or connections in past-tense inflections: what does the evidence rule out? *Trends in Cognitive Science*, 6 (11), 465–472.
- McClelland, J. L. & Rogers, T. T. (2003). The parallel distributed processing approach to semantic cognition. *Nature Reviews Neuroscience*, 4, 310–322.
- McClelland, J. L. & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception. part 1: an account of basic findings. *Psychological Review*, 88 (5), 375–407.
- McClelland, J. L. & Rumelhart, D. E. (1985). Distributed memory and the representation of general and specific information. *Journal of Experimental Psychology: General*, 114 (2), 159–188.
- McClelland, J. L. & Rumelhart, D. E. (1986). *Parallel distributed processing: explorations in the microstructure of cognition. Volume 2: psychological and biological models.* Cambridge, MA: MIT Press.
- McClelland, J. L., St. John, M. & Taraban, R. (1989). Sentence comprehension: a parallel distributed processing approach. *Language and Cognitive Processes*, 4, 287–336.
- McCloskey, M. (1991). Networks and theories: the place of connectionism in cognitive science. *Psychological Science*, 2, 387–395.
- McKevitt, P. (1996). Integration of natural language and vision processing: grounding representations. *Artificial Intelligence Review*, 10 (1-2), 7–13.
- McKevitt, P. & Guo, C. (1996). From Chinese rooms to Irish rooms: new words on visions for language. *Artificial Intelligence Review*, 10 (1-2), 49–63.
- McLeod, P., Plunkett, K. & Rolls, E. T. (1998). *Introduction to connectionist modelling of cognitive processes.* Oxford: Oxford University Press.
- McRae, K., Spivey-Knowlton, M. J. & Tanenhaus, M. K. (1998). Modeling the influence of thematic fit (and other constraints) in on-line sentence comprehension. *Journal of Memory and Language*, 38 (3), 283–312.
- Meibauer, J., Demske, U., Geilfuß-Wolfgang, J., Pafel, J., Ramers, K. H., Rothweiler, M. et al. (2002). *Einführung in die germanistische Linguistik.* Stuttgart: Metzler.
- Meini, C. & Paternoster, A. (1996). Understanding language through vision. *Artificial Intelligence Review*, 10 (1-2), 37–48.
- Mendelsohn, A. A. (2002). *Individual differences in ambiguity resolution: working memory and inhibition.* Doctoral thesis, Northeastern University Boston, MA.
- Minsky, M. & Papert, S. (1972). Research at the laboratory in vision, language and other problems of intelligence. *MIT Artificial Intelligence Memo*, 252.

- Mirman, D., Dixon, J. A. & Magnuson, J. S. (2008). Statistical and computational models of the visual world paradigm: growth curves and individual differences. *Journal of Memory and Language*, 59 (4), 475–494.
- Misyak, J. B., Christiansen, M. H. & Tomblin, J. B. (2008). Statistical learning of nonadjacencies predicts on-line processing of long-distance dependencies in natural language. In N. A. Taatgen, H. van Rijn, J. Nerbonne & L. Schomaker (Hrsg.), *Proceedings of the 31st Annual Conference of the Cognitive Science Society* (S. 177–182). Austin, TX.
- Mitchell, D. C. (1994). Sentence parsing. In M. A. Gernsbacher (Hrsg.), *Handbook of Psycholinguistics* (S. 375–405). New York: Academic Press.
- Mitkov, R. (1999). *Anaphora resolution: the state of the art* (Bericht). Wolverhampton, UK: University of Wolverhampton.
- Müller, D., Hachmann, W. & Konieczny, L. (2009). Local syntactic and semantic coherence in reading. In *Proceedings of the 15th Annual Conference on Architectures and Mechanisms for Language Processing (AMLaP)*. Barcelona, Spain.
- Morales, M. (2010). *sciplot: scientific graphing functions for factorial designs* [Software-Handbuch]. Verfügbar unter <http://CRAN.R-project.org/package=sciplot> (07.06.2010) (R package version 1.0-7)
- Morgan, E., Keller, F. & Steedman, M. (2010). A bottom-up parsing model of local coherence effects. In *Proceedings of the 32nd Annual Conference of the Cognitive Science Society* (S. 1559–1564).
- Narayanan, S. (1997). *KARMA: knowledge-based active representations for metaphor and aspect*. Doctoral thesis, Computer Science Division, EECS Department, University of California, Berkeley, USA.
- Newell, A. (1980). Physical symbol systems. *Cognitive Science*, 4, 135–183.
- Newell, A. (1990). *Unified theories of cognition*. London, UK: Harvard University Press.
- Nicol, J. L. & Swinney, D. (1989). The role of structure in coreference assignment during sentence processing. *Journal of Psycholinguistic Research*, 18 (1), 5–19.
- Niefind, F. (2009). *Der Erwerb von Anaphern in künstlichen neuronalen Netzen. Einflüsse der Grammatik auf den Erwerbsprozess*. Bachelors thesis, Universität Freiburg, Deutsches Seminar.
- Niklasson, L. F. & van Gelder, T. (1994). On being systematically connectionist. *Mind & Language*, 9, 288–302.
- Nyamapfene, A. & Ahmad, K. (2007). A multimodal model of child language acquisition at the one-word stage. In *Proceedings of the International Joint Conference on Neural Networks*. Orlando, Florida.
- Ogden, C. K. & Richards, I. A. (1923). *The meaning of meaning*. Harcourt Brace Jovanovic Publishers.
- O’Grady, W. (2008). Innateness, universal grammar, and emergentism. *Lingua*, 118, 620–631.
- Otten, M. & van Berkum, J. J. A. (2008). Discourse-based word anticipation during language processing: prediction or priming? *Discourse Processes*, 45 (6), 464–496.
- Papafragou, A. (2005). Relations between language and thought: individuation and the count/mass distinction. In C. Lefebvre & H. Cohen (Hrsg.), *Handbook of Categorization in the Cognitive Sciences* (S. 1–33). Oxford, UK: Elsevier Science.

- Papafragou, A. (2008). Space and the language–cognition interface. In P. Carruthers, S. Laurence & S. Stich (Hrsg.), *The innate mind: foundations and the future*. Oxford: Oxford University Press.
- Paradis, E., Claude, J. & Strimmer, K. (2004). APE: analyses of phylogenetics and evolution in R language. *Bioinformatics*, 20, 289–290.
- Partridge, D. (1995). Language and vision: a single perceptual mechanism? *Artificial Intelligence Review*, 9, 291–303.
- Peschl, M. (1996). Repräsentation in natürlichen und künstlichen (konnektionistischen) neuronalen Systemen. In G. Strube, B. Becker, C. Freska, U. Hahn, K. Opwis & G. Palm (Hrsg.), *Wörterbuch der Kognitionswissenschaft* (S. 579–580). Stuttgart: Klett-Cotta.
- Piattelli-Palmarini, M. (1994). Ever since language and learning: afterthoughts on the piaget-chomsky debate. *Cognition*, 50, 315–346.
- Pinker, S. (1984). *Language learnability and language development*. Cambridge, MA: Harvard University Press.
- Pinker, S. (1994). *The language instinct: the new science of language and mind*. New York: William Morrow and Company.
- Pinker, S. & Prince, A. (1988). On language and connectionism. Analysis of a parallel distributed processing model of language acquisition. *Cognition*, 28, 73–193.
- Pinker, S. & Ullman, M. T. (2002). The past-tense debate. The past and future of the past tense. *Trends in Cognitive Sciences*, 6 (11), 456–463.
- Plaut, D. C. (2002). Graded modality specific specialisation in semantics: a computational account of optic aphasia. *Cognitive Neuropsychology*, 19 (7), 603–639.
- Plaut, D. C. & McClelland, J. L. (2010). Locating object knowledge in the brain: comment on Bowers's (2009) attempt to revive the grandmother cell hypothesis. *Cognitive Neuropsychology*, 117 (1), 284–290.
- Plunkett, K. & Elman, J. L. (1997). Exercises in rethinking innateness: a handbook for connectionist simulations.
- Plunkett, K., Sinha, C., Møller, M. F. & Strandsby, O. (1992). Symbol grounding or the emergence of symbols? Vocabulary growth in children and a connectionist net. *Connection Science*, 4, 293–312.
- Pollack, J. (1990). Recursive distributed representations. *Artificial Intelligence*, 46 (1), 159–216.
- Pollard, C. & Sag, I. (1992). Anaphors in English and the scope of binding theory. *Linguistic Inquiry*, 23, 261–303.
- Pollard, C. & Sag, I. (1994). *Head driven phrase structure grammar*. Chicago: Chicago University Press.
- Pospeschill, M. (2004). *Konnektionismus und Kognition*. Stuttgart: Kohlhammer.
- Price, P., Ostendorf, M., Shattuck-Hufnagel, S. & Fong, C. (1991). The use of prosody in syntactic disambiguation. *Journal of the Acoustical Society of America*, 90, 2956–2970.
- Pulvermüller, F. (1999). Words in the brain's language. *Behavioral and Brain Sciences*, 22, 253–336.
- Pulvermüller, F., Härle, M. & Hummel, F. (2001). Walking or talking? Behavioral and neurophysiological correlates of action verb processing. *Brain and Language*, 78, 143–168.

- Putnam, H. (1975). *Mind, language and reality: philosophical papers* (Bd. 2). Cambridge, UK: Cambridge University Press.
- Pylyshyn, Z. (2003). Return of the mental image: are there really pictures in the brain? *Trends in Cognitive Sciences*, 7 (3), 113–118.
- Quine, W. V. (1960). *Word and object*. New York: Wiley.
- R Development Core Team. (2010). R: a language and environment for statistical computing [Software-Handbuch]. Vienna, Austria. Verfügbar unter <http://www.R-project.org> (06.10.2010)
- Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124, 372–422.
- Reali, F. & Christiansen, M. H. (2005). Uncovering the richness of the stimulus: structure dependence and indirect statistical evidence. *Cognitive Science*, 29, 1007–1028.
- Reali, F., Christiansen, M. H. & Monaghan, P. (2003). Phonological and distributional cues in syntax acquisition: scaling up the connectionist approach to multiple-cue integration. In R. Alterman & D. Kirsh (Hrsg.), *Proceedings of the 25th Annual Conference of the Cognitive Science Society* (S. 970–975). Mahwah, NJ: Lawrence Erlbaum Associates.
- Regier, T. (1995). A model of the human capacity for categorizing spatial relations. *Cognitive Linguistics*, 6 (1), 63–88.
- Regier, T. (1996). *The human semantic potential: spatial language and constrained connectionism*. Cambridge, MA: MIT Press.
- Regier, T. & Carlson, L. A. (2001). Grounding spatial language in perception: an empirical and computational investigation. *Journal of Experimental Psychology: General*, 130 (2), 273–298.
- Reilly, R. G. (1984). A connectionist model of some aspects of anaphor resolution. In *Proceedings of the 10th international conference on computational linguistics*. Stroudsburg, PA.
- Reilly, R. G. (1995). Sandy ideas and coloured days: some computational implications of embodiment. *Artificial Intelligence Review*, 9, 305–322.
- Reinhart, T. & Reuland, E. (1993). Reflexivity. *Linguistic Inquiry*, 24, 657–720.
- Richardson, D., Spivey, M., Barsalou, L. & McRae, K. (2003). Spatial representations activated during real-time comprehension of verbs. *Cognitive Science*, 27, 767–780.
- Robert-Ribes, J., Schwartz, J. & Escudier, P. (1995). A comparison of models for fusion of the auditory and visual sensors in speech perception. *Artificial Intelligence Review*, 9 (4), 323–346.
- Roessler, I. & Ungerer, A. (2010). *Kommentierte Formelsammlung multivariater statistischer Verfahren*. Verfügbar unter <http://www.prof-roessler.de/Dateien/Statistik/multivariat.pdf> (01.03.2010)
- Rogers, T. T. & McClelland, J. L. (2003). *Semantic cognition: a parallel distributed processing approach*. Boston, MA: MIT Press.
- Rogers, T. T. & Plaut, D. C. (2002). Connectionist perspectives on category-specific deficits. In E. Forde & G. Humphreys (Hrsg.), *Category-specificity in brain and mind*. East Sussex, UK: Psychology Press.
- Rohde, D. L. T. (1999a). *LENS: the light, efficient network simulator* (Bericht). Pittsburgh, PA: School of Computer Science, Carnegie Mellon University.

- Rohde, D. L. T. (1999b). *The simple language generator: encoding complex languages with simple grammars* (Bericht). Pittsburgh, PA: School of Computer Science, Carnegie Mellon University.
- Rohde, D. L. T. (2001). *Linger: a flexible platform for language processing experiments* (Bericht). Pittsburgh, PA: School of Computer Science, Carnegie Mellon University.
- Rohde, D. L. T. (2002). *A connectionist model of sentence comprehension and production*. Doctoral thesis, Computer Science Department, Carnegie Mellon University, Pittsburgh, PA.
- Rohde, D. L. T., Gonnerman, L. M. & Plaut, D. C. (2005). *An improved model of semantic similarity based on lexical co-occurrence*. (Unpublished manuscript)
- Rohde, D. L. T. & Plaut, D. C. (1999). Language acquisition in the absence of explicit negative evidence: how important is starting small? *Cognition*, 72, 67–109.
- Rohrer, T. (2001). Pragmatism, ideology and embodiment: William James and the philosophical foundations of cognitive linguistics. In R. Dirven, B. W. Hawkins & E. Sandikcioglu (Hrsg.), *Language and ideology: Cognitive theoretical approaches* (S. 49–82). Amsterdam: John Benjamins.
- Rohrer, T. (2005). Image schemata in the brain. In B. Hampe & J. Grady (Hrsg.), *From perception to meaning: image schemas in Cognitive Linguistics* (S. 165–196). Berlin: Mouton de Gruyter.
- Rohrer, T. (2006). Three dogmas of embodiment: Cognitive Linguistics as a Cognitive Science. In R. Dirven, G. Kristiansen & M. Achard (Hrsg.), *Cognitive Linguistics: current applications and future applications* (S. 119–146). Berlin: Mouton de Gruyter.
- Rohrer, T. (2007). The body in space: embodiment, experientialism and linguistic conceptualization. In T. Ziemke, J. Zlatev, R. Frank & R. Dirven (Hrsg.), *Body, language and mind* (Bd. 1, S. 339–378). Berlin: Mouton de Gruyter.
- Rosch, E. (1977). Human categorization. In N. Warren (Hrsg.), *Studies in cross-cultural psychology* (S. 1–47). London, UK: Academic Press.
- Rosch, E. & Lloyd, B. B. (1978). *Cognition and categorization*. Oxford, UK: Lawrence Erlbaum.
- Rosenblatt, F. (1958). The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological Review*, 65, 386–408.
- Roy, D. (1999). Learning words from sights and sounds: a computational model [Doctoral thesis].
- Rueckl, J. G., Cave, K. R. & Kosslyn, S. M. (1989). Why are „what“ and „where“ processed by separate cortical visual systems? A computational investigation. *Journal of Cognitive Neuroscience*, 1 (2), 171–186.
- Rumelhart, D. E. (1989). The architecture of mind: a connectionist approach. In M. I. Posner (Hrsg.), *Foundations of Cognitive Science* (S. 133–160). MIT Press.
- Rumelhart, D. E. (1990). Brain style computation: learning and generalization. In S. F. Zornetzer, J. L. Davis & C. Lau (Hrsg.), *An introduction to neural and electronic networks* (S. 405–420). San Diego, CA: Academic Press.
- Rumelhart, D. E., Hinton, G. E. & Williams, R. J. (1986). Learning internal representations by error propagation. In D. E. Rumelhart & J. L. McClelland (Hrsg.), *Parallel distributed processing: explorations in the microstructure of cognition. Volume 1: foundations* (S. 318–362). Cambridge, MA: MIT Press.

- Rumelhart, D. E. & McClelland, J. L. (1986a). On learning the past tense of English verbs. In D. E. Rumelhart & J. L. McClelland (Hrsg.), *Parallel distributed processing: explorations in the microstructure of cognition. Volume 1: foundations* (S. 216–271). Cambridge, MA: MIT Press.
- Rumelhart, D. E. & McClelland, J. L. (1986b). *Parallel distributed processing: explorations in the microstructure of cognition. Volume 1: foundations*. Cambridge, MA: MIT Press.
- Rumelhart, D. E., Smolensky, P., McClelland, J. L. & Hinton, G. E. (1986). Schemata and sequential thought processes in PDP models. In J. L. McClelland & D. E. Rumelhart (Hrsg.), *Parallel distributed processing: explorations in the microstructure of cognition. volume 2: psychological and biological models* (S. 7–57). Cambridge, MA: MIT Press.
- Runner, J. T., Sussman, R. S. & Tanenhaus, M. K. (2003). Assignment of reference to reflexives and pronouns in picture noun phrases: evidence from eye movements. *Cognition*, 89 (1), B1–B13.
- Runner, J. T., Sussman, R. S. & Tanenhaus, M. K. (2006). Processing reflexives and pronouns in picture noun phrases. *Cognitive Science*, 30, 193–241.
- Saffran, J. R. (2001). The use of predictive dependencies in language learning. *Journal of Memory and Language*, 44, 493–515.
- Saffran, J. R. (2002). Constraints on statistical language learning. *Journal of Memory and Language*, 47, 172–196.
- Saffran, J. R. (2003). Statistical language learning: mechanisms and constraints. *Current Directions in Psychological Science*, 12 (4), 110–114.
- Sales, N. J., Evans, R. G. & Aleksander, I. (1996). Successful naive representation grounding. *Artificial Intelligence Review*, 10 (1-2), 83–102.
- Santiago, J., Lupianez, J., Pérez, E. & Funes, M. (2007). Time (also) flies from left to right. *Psychonomic Bulletin & Review*, 14 (3), 512–516.
- Schank, R. C. & Fano, A. (1995). Memory and expectations in learning, language, and visual understanding. *Artificial Intelligence Review*, 9, 261–271.
- Schmitt, R. (2001). Metaphern in der Psychologie – eine Skizze. *Journal für Psychologie*, 9 (4), 3–15.
- Schwarz, M. & Chur, J. (2004). *Semantik. Ein Arbeitsbuch*. Tübingen: Narr.
- Schwarzkopf, S. (2008). *Mentale Simulation von Perspektive bei der Satzverarbeitung. Eine experimentelle Studie zur Embodied Cognition*. Masters thesis, Universität Freiburg, Deutsches Seminar.
- Schwarzkopf, S., Müller, D., Weldle, H. & Konieczny, L. (2008). Grammatical person controls perspective taking in embodied simulation semantics. In *Proceedings of the 14th Annual Conference on Architectures and Mechanisms for Language Processing (AMLaP)*. Cambridge, UK.
- Schwarzkopf, S., Weldle, H., Müller, D. & Konieczny, L. (2011, accepted). Mental simulation of spatial perspective during sentence comprehension. In *Proceedings of the 33rd Annual Conference of the Cognitive Science Society*. Boston, USA.
- Scorrolli, C. & Borghi, A. M. (2007). Sentence comprehension and action: effector specific modulation of the motor system. *Brain Research*, 1130, 119–124.
- Searle, J. (1980). Minds, brains and programs. *Behavioural and Brain Sciences*, 3, 417–457.

- Sedivy, J. C., Tanenhaus, M. K., Chambers, C. G. & Carlson, G. N. (1999). Achieving incremental semantic interpretation through contextual representation. *Cognition*, 71, 109–147.
- Seidenberg, M. S. & MacDonald, M. C. (1999). A probabilistic constraints approach to language acquisition and processing. *Cognitive Science*, 23, 569–588.
- Seidenberg, M. S. & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review*, 96, 523–568.
- Shapiro, L. (2007). The embodied cognition research programme. *Cognition*, 2, 338–346.
- Shultz, T. R., Buckingham, D. & Oshima-Takane, Y. (1994). A connectionist model of the learning of personal pronouns in English. In S. J. Hanson, T. Petsche, M. Kearns & R. L. Rivest (Hrsg.), *Computational learning theory and natural learning systems. volume 2: intersection between theory and experiment* (S. 347–362). Cambridge, MA: MIT Press.
- Simon, H. & Wallach, D. (1999). Cognitive modelling in perspective. *Kognitionswissenschaft*, 8, 1–4.
- Siskind, J. M. (1995). Grounding language in perception. *Artificial Intelligence Review*, 8, 371–391.
- Siskind, J. M. (1996). A computational study of cross-situational techniques for learning word-to-meaning mappings. *Cognition*, 61, 39–91.
- Slobin, D. (1966). Grammatical transformations and sentence comprehension in childhood and adulthood. *Journal of Verbal Learning and Verbal Behavior*, 5, 219–227.
- Smolensky, P. (1988). On the proper treatment of connectionism. *Behavioral and Brain Sciences*, 11 (1), 1–74.
- Smolensky, P. (1990). Tensor product variable binding and the representation of symbolic structure in connectionist systems. *Artificial Intelligence*, 46 (1-2), 159–216.
- Smolensky, P. (1999). Grammar-based connectionist approaches to language. *Cognitive Science: A Multidisciplinary Journal*, 23, 589–613.
- Srihari, R. K. (1995). Computational models for integrating linguistic and visual information: a survey. *Artificial Intelligence Review*, 8, 349–369.
- Stanfield, R. & Zwaan, R. A. (2001). The effect of implied orientation derived from verbal context on picture recognition. *Psychological Science*, 12, 153–156.
- StatSoft, I. (2010). Electronic statistics textbook [Software-Handbuch]. Tulsa, OK. Verfügbar unter <http://www.statsoft.com/textbook/> (12.02.2010)
- Steedman, M. (1999). Connectionist sentence processing in perspective. *Cognitive Science*, 23 (4), 615–634.
- Steels, L. (2003). Evolving grounded communication for robots. *Trends in Cognitive Science*, 7 (7), 308–312.
- Steels, L., Kaplan, F., McIntyre, A. & van Looveren, J. (2002). Crucial factors in the origins of word-meaning. In A. Wray (Hrsg.), *The transition to language* (S. 252–271). Oxford, UK: Oxford University Press.
- Steinhauer, K., Alter, K. & Friederici, A. (1999). Brain potentials indicate immediate use of prosodic cues in natural speech processing. *Nature Neuroscience*, 2, 191–196.
- Stephan, A. (1999). Varieties of Emergentism. *Evolution and Cognition*, 5 (1), 49–59.
- St. John, M. F. (1992). The story gestalt: a model of knowledge-intensive processes in text comprehension. *Cognitive Science*, 16, 271–306.

- St. John, M. F. & McClelland, J. L. (1990). Learning and applying contextual constraints in sentence comprehension. *Artificial Intelligence*, 46, 217–457.
- St. John, M. F. & McClelland, J. L. (1992). Parallel constraint satisfaction as a comprehension mechanism. In R. G. Reilly & N. E. Sharkey (Hrsg.), *Connectionist approaches to natural language processing* (S. 97–136). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Stolcke, A. (1995). An efficient probabilistic context-free parsing algorithm that computes prefix probabilities. *Computational Linguistics*, 21 (2), 165–201.
- Strube, G. (1998). Modelling motivation and action control in cognitive systems. In U. Schmid, J. Krems & F. Wysotzki (Hrsg.), *Mind modeling* (S. 89–108). Lengerich: Pabst Science Publishers.
- Sturt, P. (2003). The time-course of the application of binding constraints in reference resolution. *Journal of Memory and Language*, 48, 542–562.
- Sturt, P., Pickering, M. J. & Crocker, M. W. (1999). Structural change and reanalysis difficulty in language comprehension. *Journal of Memory and Language*, 40, 136–150.
- Sun, R. & Bookman, L. A. (1994). *Computational architectures integrating neural and symbolic processes: a perspective on the state of the art*. Boston, MA: Kluwer.
- Szagan, G. (2000). *Sprachentwicklung beim Kind* (Unveränderter Nachdruck der 6., vollst. überarb. Ausgabe Aufl., Bd. 62). Weinheim: Beltz.
- Taatgen, N. & Anderson, J. (2008). Constraints in cognitive architectures. In R. Sun (Hrsg.), *Cambridge Handbook of Computational Psychology* (S. 170–185). New York: Cambridge University Press.
- Tabor, W., Galantucci, B. & Richardson, D. (2004). Effects of merely local syntactic coherence on sentence processing. *Journal of Memory and Language*, 50 (4), 355–370.
- Tabor, W. & Hutchins, S. (2004). Evidence for self-organized sentence processing: digging-in effects. *Journal of Experimental Psychology: Learning, Memory & Cognition*, 30, 431–450.
- Tabor, W., Juliano, C. & Tanenhaus, M. K. (1997). Parsing in a dynamical system: an attractor-based account of the interaction of lexical and structural constraints in sentence processing. *Language and Cognitive Processes*, 12, 211–271.
- Tabor, W. & Tanenhaus, M. K. (1999). Dynamical models of sentence processing. *Cognitive Science*, 23 (4), 491–515.
- Tanenhaus, M. K., Frank, A., Salverda, A. P., Jaeger, F. & Masharov, M. (2008). The art of the state: mixed effects regression modeling in the visual world. In *Proceedings of the 21st CUNY Conference on Human Sentence Processing*. University of North Carolina at Chapel Hill.
- Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M. & Sedivy, J. E. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, 268, 632–634.
- Tanenhaus, M. K. & Trueswell, J. (2006). Eye movements and spoken language comprehension. In M. Traxler & M. A. Gernsbacher (Hrsg.), *Handbook of Psycholinguistics* (2. Aufl., S. 863–900). San Diego: Academic Press.
- Tettamanti, M., Buccino, G., Saccuman, M. C., Rizzolatti, G., Cappa, S. F. & Perani, D. (2005). Listening to action-related sentences activates fronto-parietal motor circuits.

- Journal of Cognitive Neuroscience*, 17 (2), 273–281.
- Thompson, L. A. & Ogden, W. C. (1995). Visible speech improves human language understanding: implications for speech processing systems. *Artificial Intelligence Review*, 9, 347–358.
- Tomasello, M. (2005). *Constructing a language. A usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.
- Tomasello, M. (2006). Acquiring linguistic constructions. In D. Kuhn, R. S. Siegler, W. Damon & R. M. Lerner (Hrsg.), *Handbook of Child Psychology. Volume 2: Cognition, Perception and Language* (6. Aufl., S. 255–298). Hoboken, NJ: Wiley Publishers.
- Touretzky, D. S. (1991). Connectionism and compositional semantics. In J. A. Barnden & J. B. Pollack (Hrsg.), *High-level connectionist models. Advances in connectionist & neural computation theory* (S. 17–31). Hillsdale, NJ: Erlbaum.
- Treiman, R., Clifton Jr., C., Meyer, A. S. & Wurm, L. H. (2003). Language comprehension and production. In A. F. Healy & R. W. Proctor (Hrsg.), *Comprehensive Handbook of Psychology. Volume 4: Experimental Psychology* (S. 527–548). New York: Wiley Publishers.
- Udiljak, B. (2010). *Autoassoziative Prototypenerzeugung für den visuellen Input eines konnektionistischen 2-Wege-Modells*. Diploma thesis, Universität Freiburg, Institut für Informatik.
- Ullman, S. (1996). *High-level vision. Object recognition and visual cognition*. Cambridge, MA: MIT Press.
- Van der Velde, F., Van der Voort van der Kleij, G. T. & De Kamps, M. (2004). Lack of combinatorial productivity in language processing with simple recurrent networks. *Connection Science*, 16, 21–46.
- van Dyke, J. (2007). Interference effects from grammatically unavailable constituents during sentence parsing. *Journal of Experimental Psychology*, 33 (2), 407–430.
- van Dyke, J. & Lewis, R. L. (2003). Distinguishing effects of structure and decay on attachment and repair: a cue-based parsing account of recovery from misanalyzed ambiguities. *Journal of Memory and Language*, 49 (2), 285–316.
- van Everbroeck, E. (2003). Language type frequency and learnability from a connectionist perspective. *Linguistic Typology*, 7, 1–50.
- van Gelder, T. (1990). Compositionality: a connectionist variation on a classical theme. *Cognitive Science: A Multidisciplinary Journal*, 14 (3), 355–384.
- van Gelder, T. (1992). Defining „distributed representation“. *Connection Science*, 4 (3-4), 175–191.
- van Gelder, T. (1998). The dynamical hypothesis in cognitive science. *Behavioral and Brain Sciences*, 21, 615–665.
- van Gelder, T. & Port, R. F. (1995). It's about time: an overview of the dynamical approach to cognition. In R. Port & T. van Gelder (Hrsg.), *Mind as motion: explorations in the dynamics of cognition* (S. 1–43). Cambridge, MA: MIT Press.
- van Gompel, R. P. G. & Majid, A. (2004). Antecedent frequency effects during the processing of pronouns. *Cognition*, 90 (3), 255–264.
- Vosse, T. & Kempen, G. (2000). Syntactic structure assembly in human parsing: a computational model based on competitive inhibition and a lexicalist grammar. *Cognition*, 75 (2), 105–43.

- Vosse, T. & Kempen, G. (2009). The unification space implemented as a localist neural net: predictions and error-tolerance in a constraint-based parser. *Cognitive Neurodynamics*, 3, 331–346.
- Waskan, J. A. (2001). A critique of connectionist semantics. *Connection Science*, 13 (3), 277–292.
- Waters, G. S. & Caplan, D. (1996). The capacity theory of sentence comprehension: critique of Just and Carpenter (1992). *Psychological Review*, 103, 761–772.
- Weckerly, J. & Elman, J. (1992). A PDP approach to processing center-embedded sentences. In *Proceedings of the 14th Annual Conference of the Cognitive Science Society*. Hillsdale, NJ: Erlbaum.
- Weldle, H. (2005). *Konnektionistische Modellierung syntaktischen Sprachwandels*. Masters thesis, Universität Freiburg, Deutsches Seminar.
- Weldle, H., Baumann, P., Müller, D., Wolfer, S. A. & Konieczny, L. (2009). Connectionist modeling of syntax and anaphora resolution. In *Proceedings of the 22nd Annual Meeting of the CUNY Conference on Human Sentence Processing*. Davis, USA.
- Weldle, H., Konieczny, L., Müller, D., Wolfer, S. & Baumann, P. (2009a). Connectionist modeling of situated language processing: language and meaning acquisition from an Embodiment perspective. In N. Taatgen & H. vanRijn (Hrsg.), *Proceedings of the 31st Annual Meeting of the Cognitive Science Society* (S. 827–832). Austin, TX.
- Weldle, H., Konieczny, L., Müller, D., Wolfer, S. A. & Baumann, P. (2009b). Language acquisition and processing in integrative connectionist models. In *Proceedings of the Annual Meeting on Embodied and Situated Language Processing*. Rotterdam, NL.
- Weldle, H., Konieczny, L., Müller, D., Wolfer, S. A. & Baumann, P. (2010a). Emergence of language comprehension from multiple task integration – a connectionist investigation. In *Proceedings of the 23rd Annual Meeting of the CUNY Conference on Human Sentence Processing*. New York.
- Weldle, H., Konieczny, L., Müller, D., Wolfer, S. A. & Baumann, P. (2010b). Grounding und Emergenz von Spracherwerb und Sprachverarbeitung: Ein konnektionistischer Ansatz. In C. Frings, A. Mecklinger, D. Wentura & H. Zimmer (Hrsg.), *Beiträge zur 52. Tagung experimentell arbeitender Psychologen*. Saarbrücken.
- Weldle, H., Konieczny, L., Wolfer, S. A., Baumann, P. & Müller, D. (2010). Local syntactic coherences violate structural binding domains in on-line anaphora resolution. In *Proceedings of the 23d Annual Meeting of the CUNY Conference on Human Sentence Processing*. New York.
- Wells, J. B., Christiansen, M. H., Race, D. S., Acheson, D. J. & MacDonald, M. C. (2009). Experience and sentence processing: statistical learning and relative clause comprehension. *Cognitive Psychology*, 58, 250–271.
- Wennekers, T., Garagnani, M. & Pulvermüller, F. (2006). Language models based on hebbian cell assemblies. *Journal of Physiology*, 100, 16–30.
- Werbos, P. J. (1974). *Beyond regression: New tools for prediction and analysis in the behavioral sciences*. Doctoral thesis, Harvard University.
- Werbos, P. J. (1990). Backpropagation through time: what it does and how to do it. In *Proceedings of the IEEE* (Bd. 78, S. 1550–1560).
- Widrow, G. & Hoff, M. E. (1960). Adaptive switching circuits. *WESCON Convention Record*, 4, 96–104.

- Wiedenbeck, M. & Züll, C. (2001). Klassifikation mit Clusteranalyse: Grundlegende Techniken hierarchischer und K-means-Verfahren. *ZUMA How-to-Reihe* (10).
- Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin & Review*, 9 (4), 625–636.
- Wolfer, S. (2007). *Semantische Verarbeitung in konnektionistischen Netzwerken*. Masters thesis, Universität Freiburg, Deutsches Seminar.
- Yu, C. (2006). Learning syntax-semantics mappings to bootstrap word learning. In *Proceedings of the 28th Annual Meeting of the Cognitive Science Society*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Yu, C. & Ballard, D. H. (2003). *A computational model of embodied language learning* (Bericht). Rochester, New York: Computer Science Department, University of Rochester.
- Yu, C. & Ballard, D. H. (2004). On the integration of grounding language and learning objects. In *Proceedings of the 19th national conference on Artificial Intelligence (AAAI)* (S. 488–493). San Jose, CA.
- Yu, C. & Smith, L. B. (2006). Statistical cross-situational learning to build word-to-world mappings. In *Proceedings of the 28th Annual Meeting of the Cognitive Science Society*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Zwaan, R. A. (2004). The immersed experiencer: toward an embodied theory of language comprehension. In B. H. Ross (Hrsg.), *The psychology of learning and motivation* (Bd. 44, S. 35–62). New York: Academic Press.
- Zwaan, R. A. & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychology Bulletin*, 123, 162–185.
- Zwaan, R. A., Stanfield, R. A. & Yaxley, R. H. (2002). Do language comprehenders routinely represent the shapes of objects. *Psychological Science*, 13 (2), 168–171.
- Zwaan, R. A. & Taylor, L. J. (2006). Seeing, acting, understanding: motor resonance in language comprehension. *Journal of Experimental Psychology: General*, 135 (1), 1–11.
- Zwaan, R. A. & Yaxley, R. H. (2003). Spatial iconicity affects semantic-relatedness judgments. *Psychonomic Bulletin & Review*, 10, 954–958.